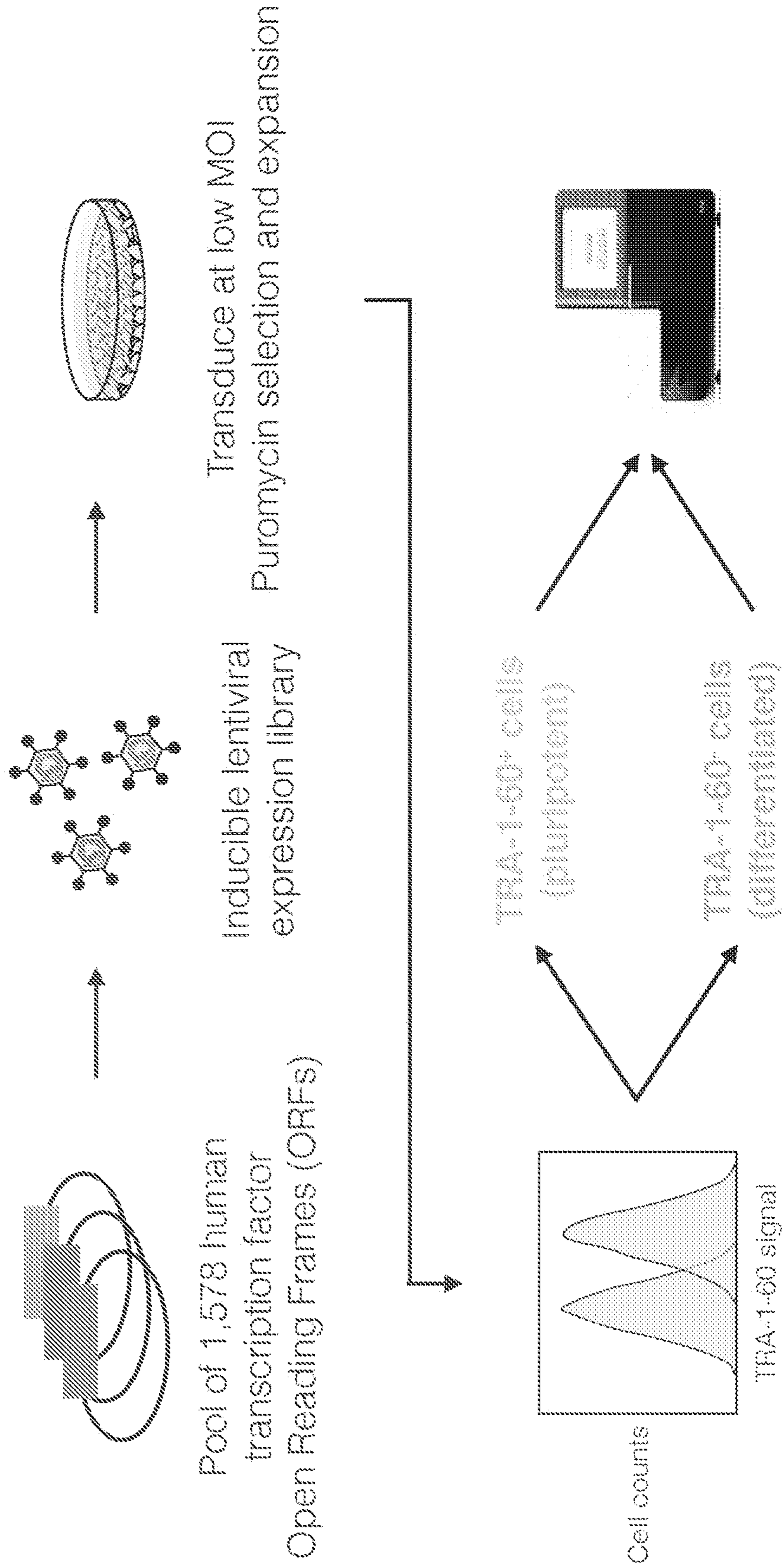


(19) **United States**(12) **Patent Application Publication**  
**Ng et al.**(10) **Pub. No.: US 2024/0209321 A1**(43) **Pub. Date: Jun. 27, 2024**(54) **TRANSCRIPTION FACTORS  
CONTROLLING DIFFERENTIATION OF  
STEM CELLS***C12N 5/0775* (2006.01)*C12N 5/0797* (2006.01)(71) Applicant: **President and Fellows of Harvard  
College, Cambridge, MA (US)**(72) Inventors: **Hon Man Alex Ng, Cambridge, MA  
(US); George M. Church, Cambridge,  
MA (US); Volker Busskamp,  
Cambridge, MA (US)**(73) Assignee: **President and Fellows of Harvard  
College, Cambridge, MA (US)**(52) **U.S. Cl.**CPC ..... *C12N 5/069* (2013.01); *C12N 5/0605*  
(2013.01); *C12N 5/0623* (2013.01); *C12N*  
*5/0626* (2013.01); *C12N 5/0658* (2013.01);  
*C12N 5/0662* (2013.01); *C12N 5/0686*  
(2013.01); *C12N 5/0696* (2013.01); *C12N*  
*2501/60* (2013.01); *C12N 2503/02* (2013.01);  
*C12N 2506/45* (2013.01); *C12N 2510/00*  
(2013.01)(21) Appl. No.: **18/496,419**(22) Filed: **Oct. 27, 2023****Related U.S. Application Data**(63) Continuation of application No. 16/332,350, filed on  
Mar. 11, 2019, now Pat. No. 11,845,960, filed as  
application No. PCT/US17/51122 on Sep. 12, 2017.(60) Provisional application No. 62/393,324, filed on Sep.  
12, 2016, provisional application No. 62/411,728,  
filed on Oct. 24, 2016, provisional application No.  
62/492,552, filed on May 1, 2017, provisional appli-  
cation No. 62/517,307, filed on Jun. 9, 2017.**Publication Classification**(51) **Int. Cl.***C12N 5/071* (2006.01)*C12N 5/073* (2006.01)*C12N 5/074* (2006.01)*C12N 5/077* (2006.01)(57) **ABSTRACT**

Forced expression of a handful of transcription factors (TFs) can induce conversions between cell identities; however, the extent to which TFs can alter cell identity has not been systematically assessed. Here, we assembled a human TFome, a comprehensive expression library of 1,578 human TF clones with full coverage of the major TF families. By systematically screening the human TFome, we identified 77 individual TFs that induce loss of human-induced-pluripotent-stem-cell (hiPSC) identity, suggesting a pervasive ability for TFs to alter cell identity. Using large-scale computational cell type classification trained on thousands of tissue expression profiles, we identified cell types generated by these TFs with high efficiency and speed, without additional selections or mechanical perturbations. TF expression in adult human tissues only correlated with some of the cell lineage generated, suggesting more complexity than observation studies can explain.

**Figure 1. Workflow for screening the human TFome for loss of pluripotency in hiPSCs**



Doxycycline induction, then sort TPA-1-60 populations

**Figure 2. General strategy for high-throughput screening for cell type conversion**

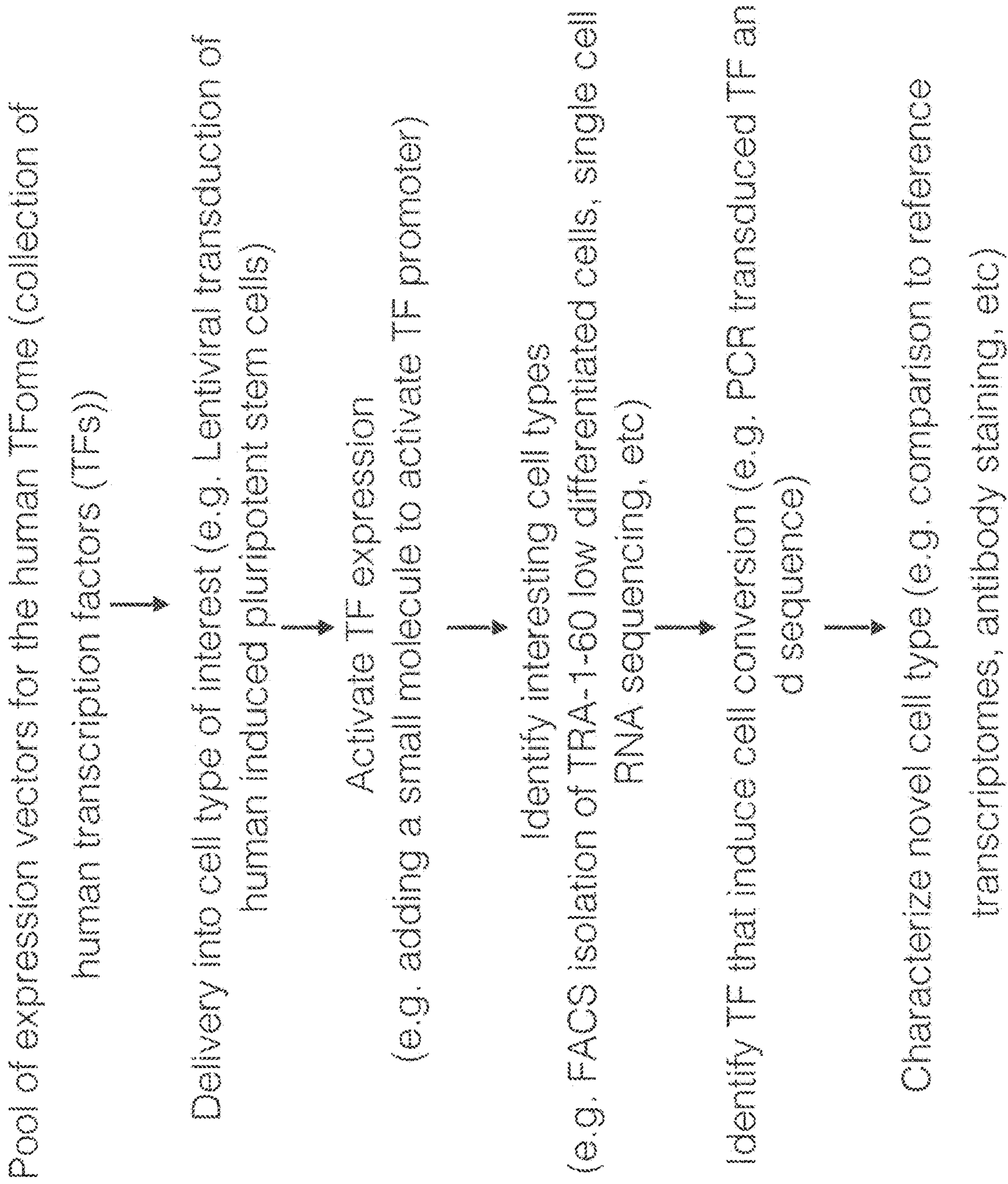
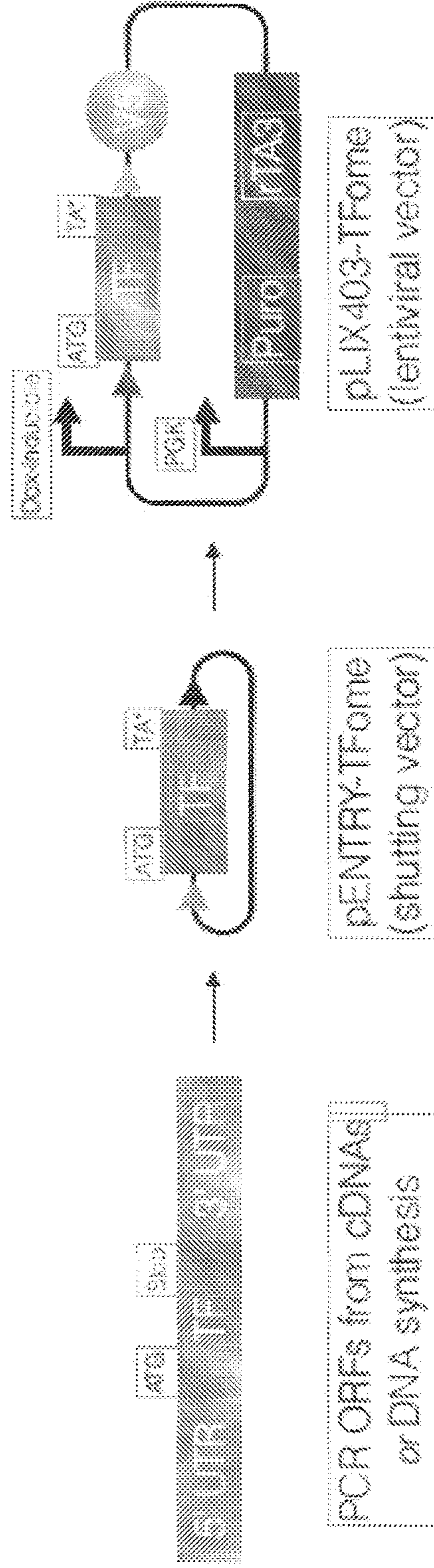


Figure 3. The Human TFome expression library



Size distribution of TFs

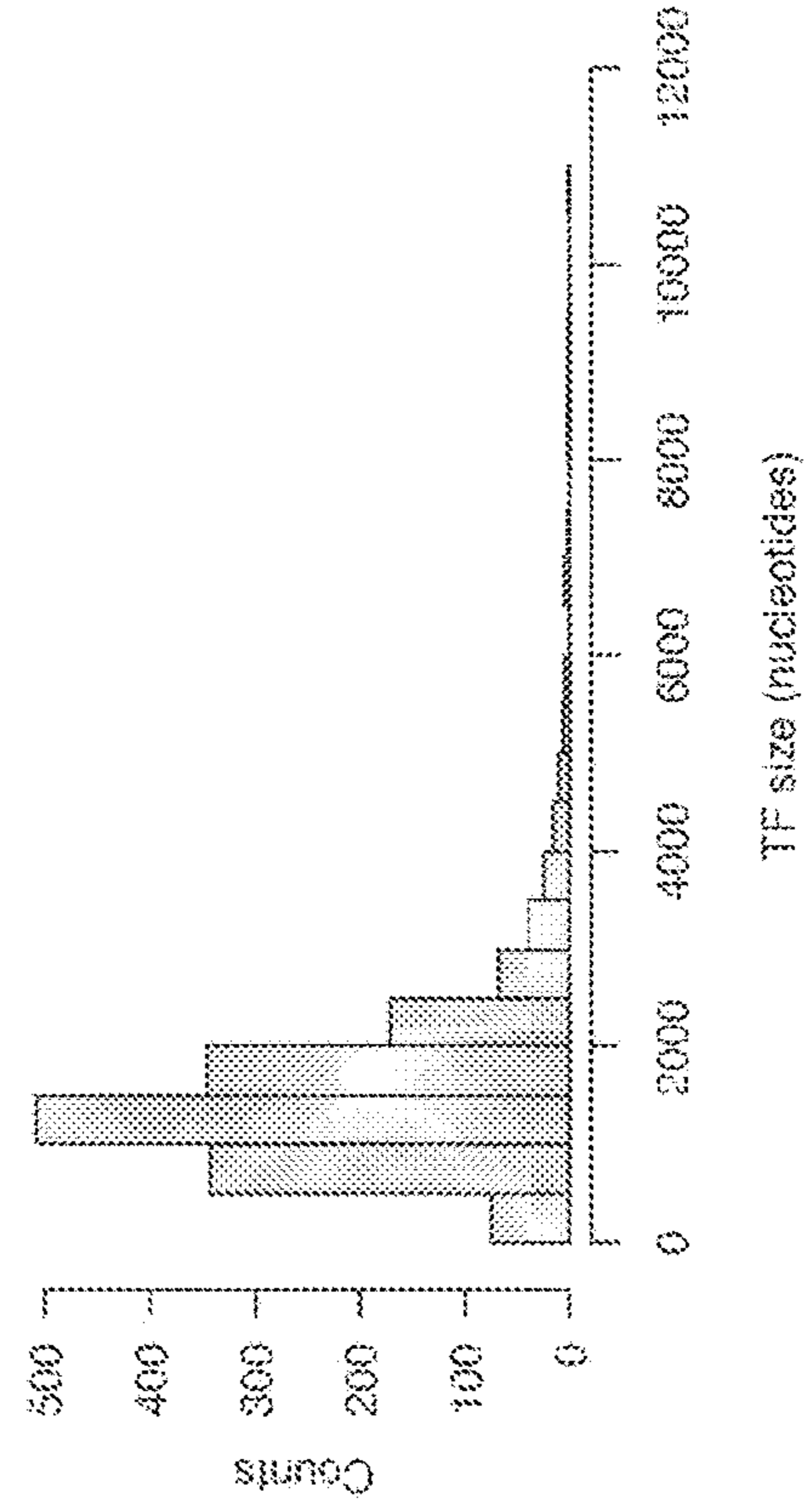
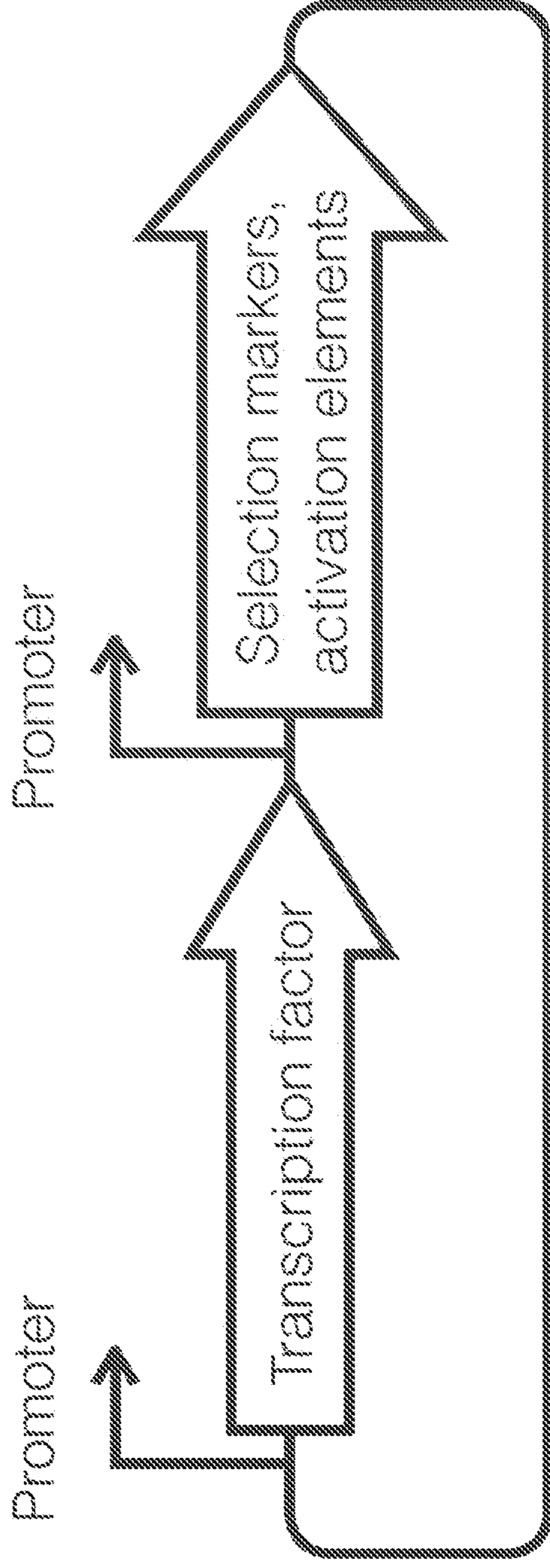


Figure 4. Expression vectors for delivery of human Trome



Example promoters:

Constitutive expression (e.g. EF1alpha promoter)

Inducible expression (e.g. small molecule, heat-shock, light) Cell type-

dependent expression (e.g. neuronal promoter)

Example delivery vectors:

Transient expression (e.g. plasmid with no mammalian integration ability)

Viral vectors (e.g. lentiviral, retroviral, AAV)

Transposon-based (e.g. PiggyBac)

Figure 5. Hits from high-throughput screening for stem cell differentiation [loss of stem cell identity (TRA-1-60) in human induced pluripotent stem cells]

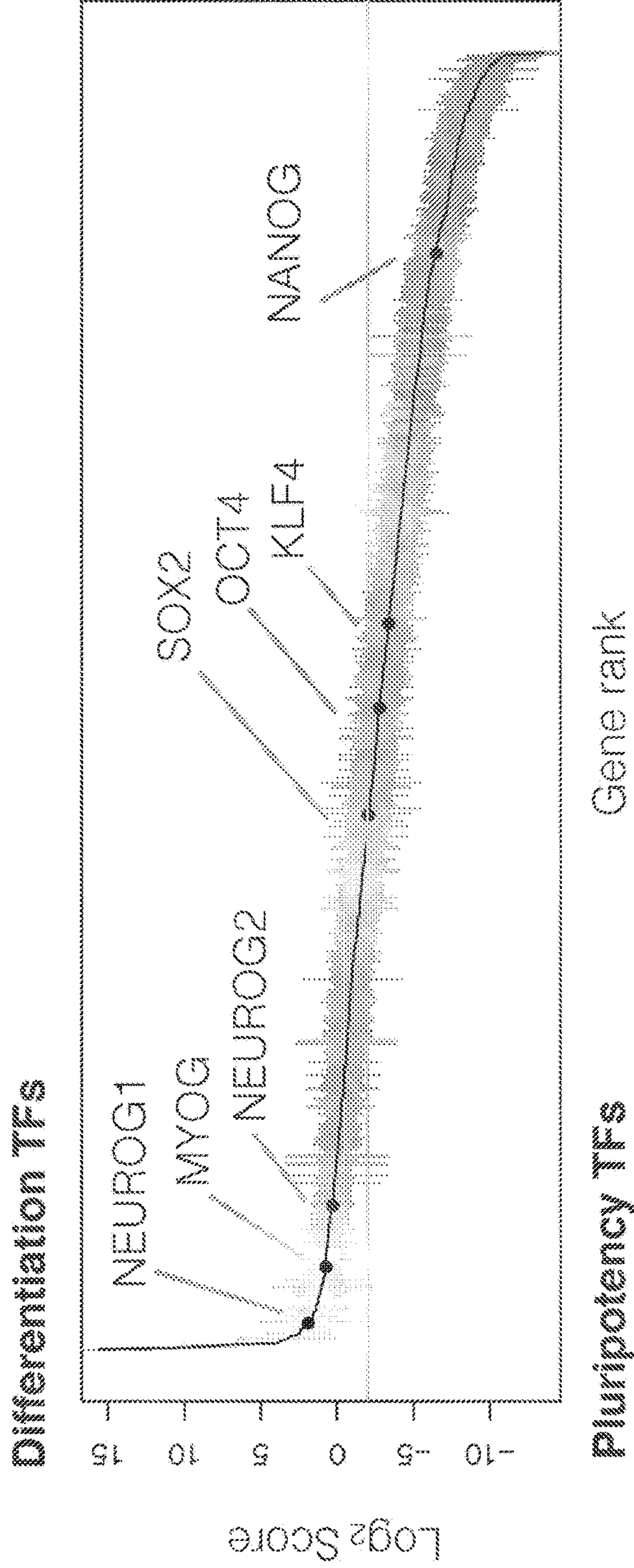
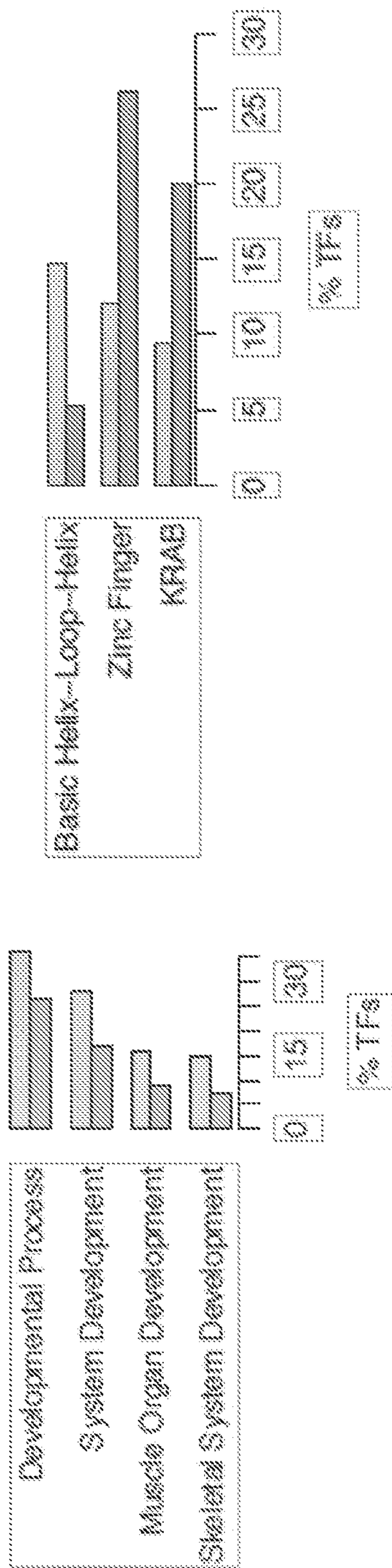
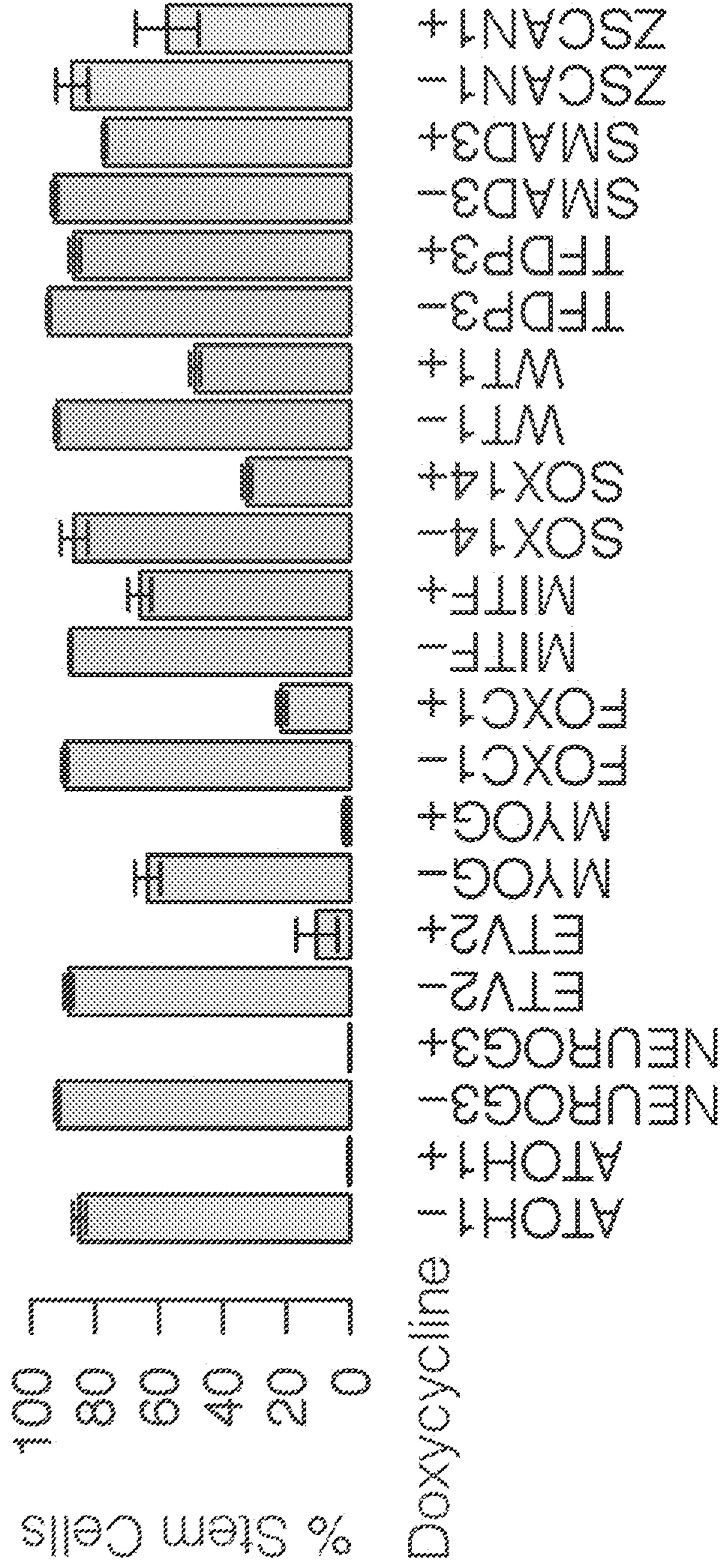


Figure 6. Hits are enriched for developmental genes and protein domains



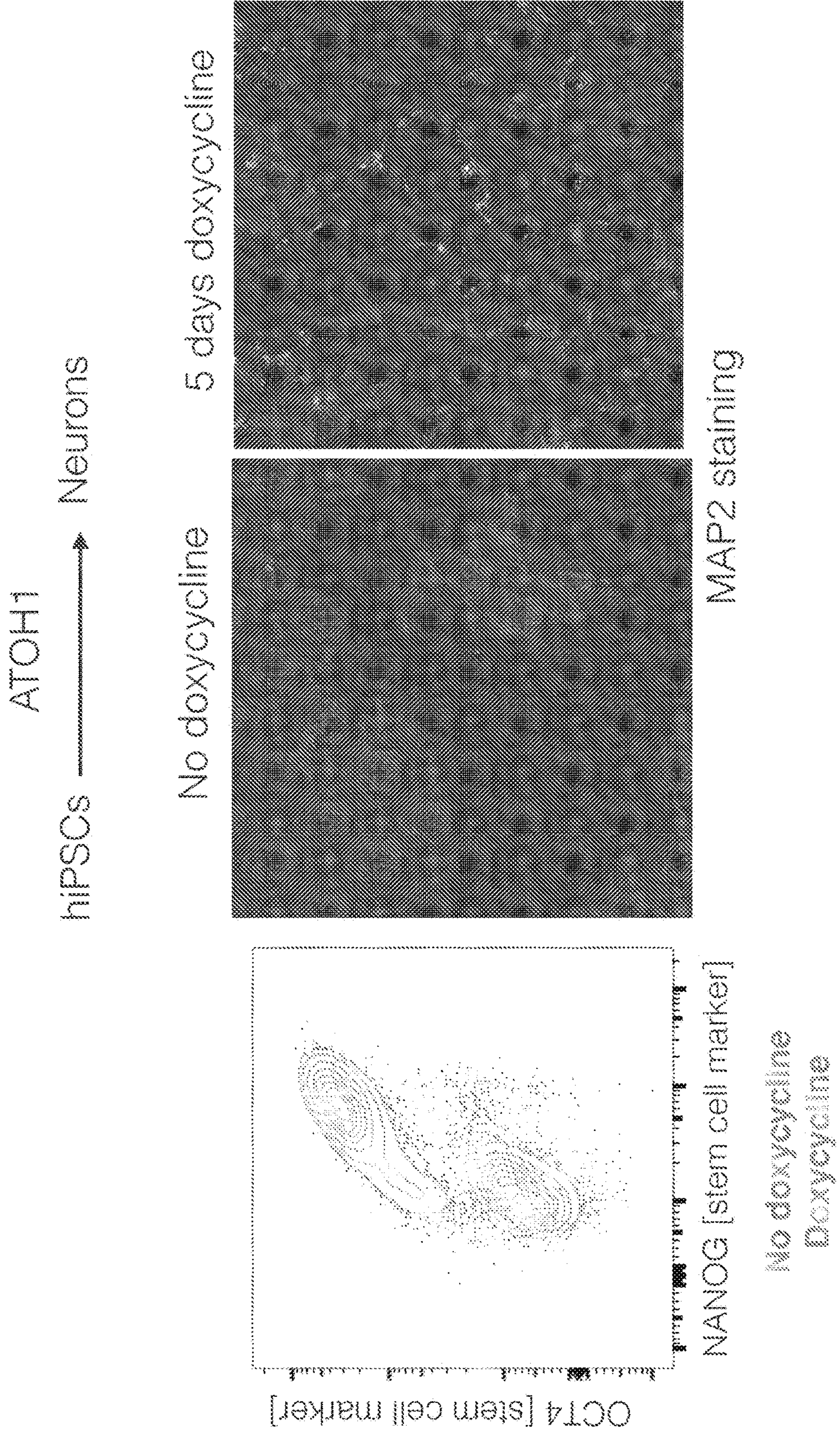
Differentiation TFs  
All TFs

Figure 7. Selected transcription factors that induce stem cell differentiation

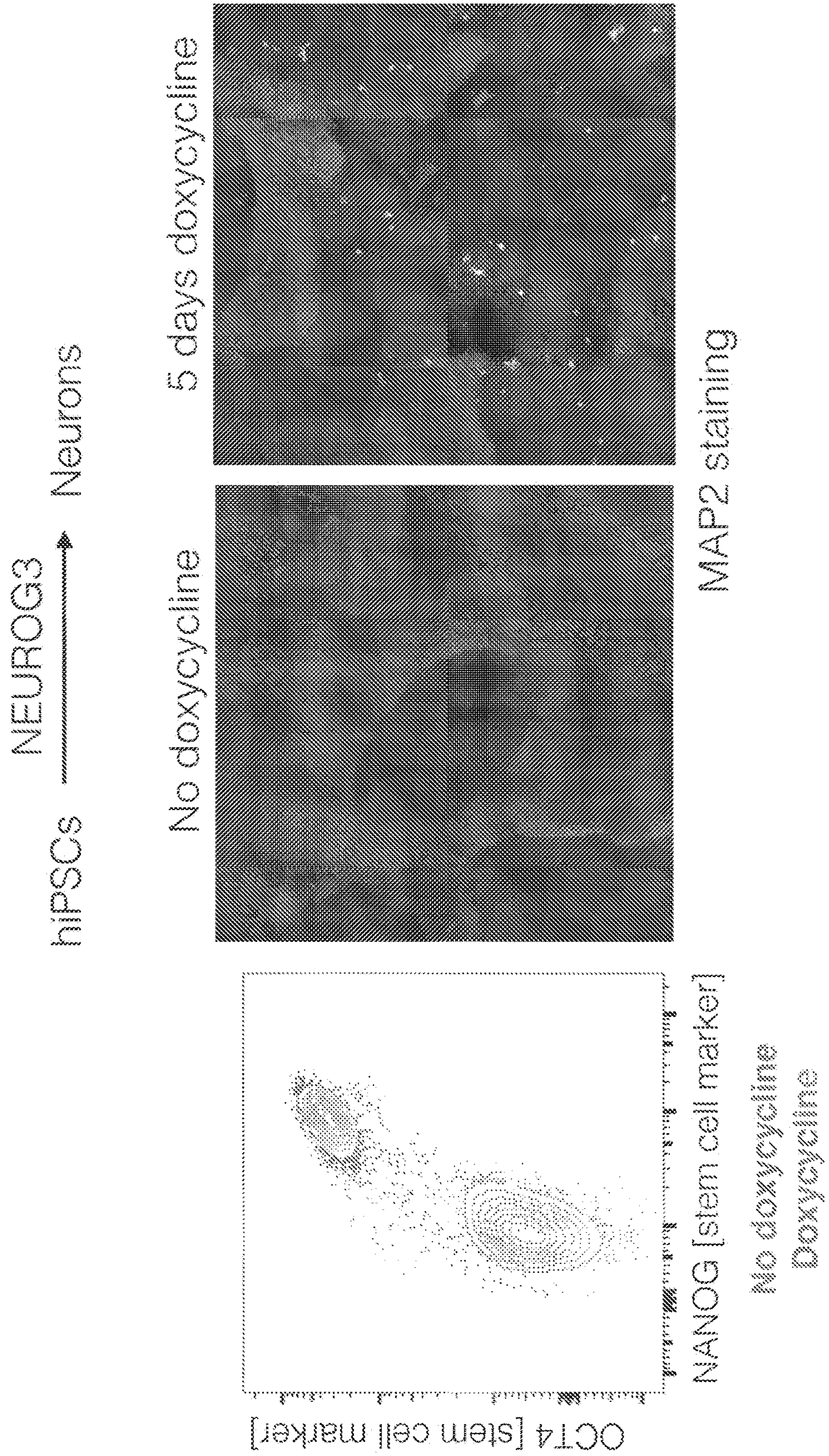




**Figure 8. ATOH1 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

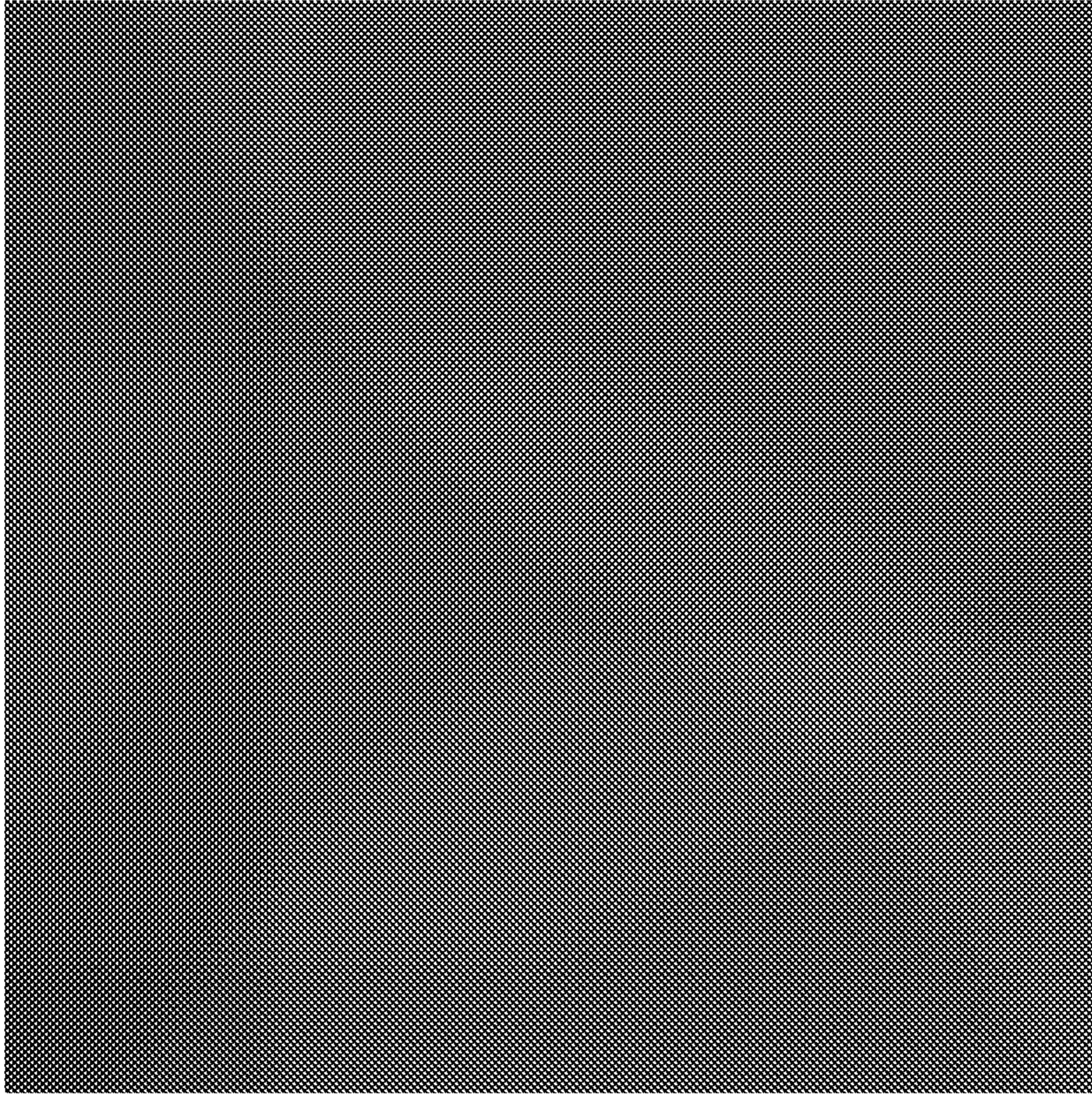


**Figure 9. NEUROG3 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

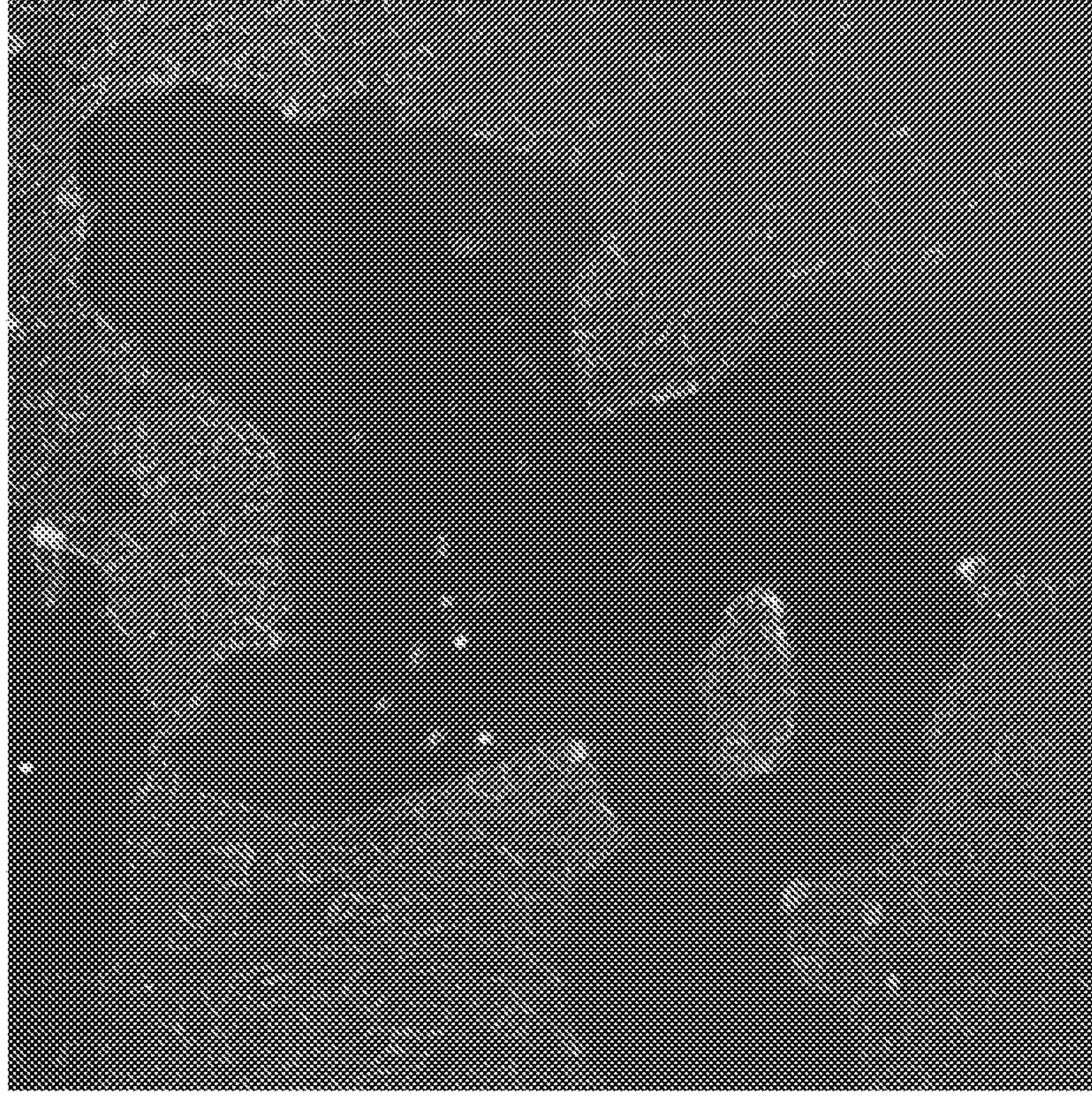


**Figure 10. ETV2 induces endothelial differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

No doxycycline



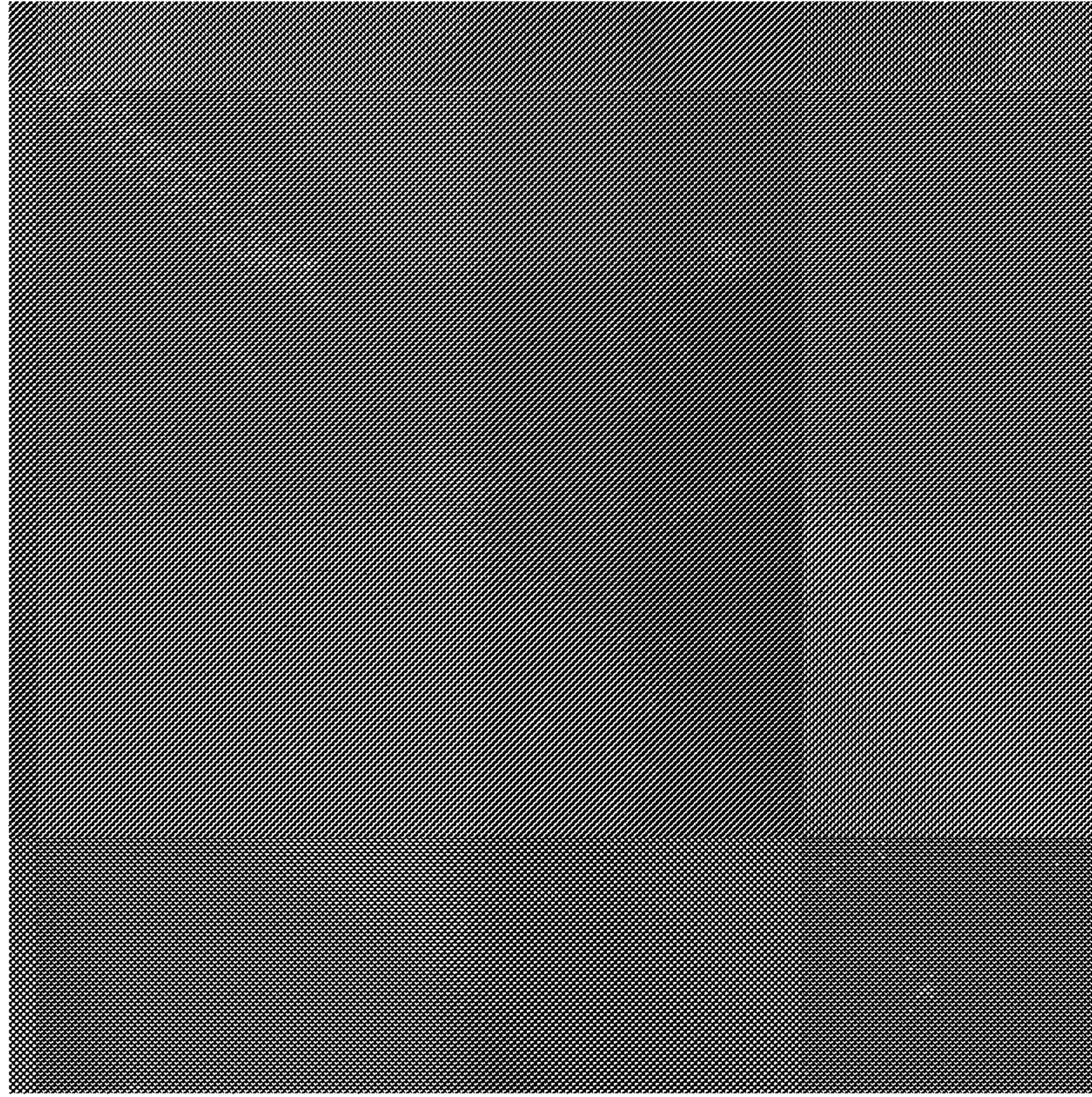
6 days doxycycline



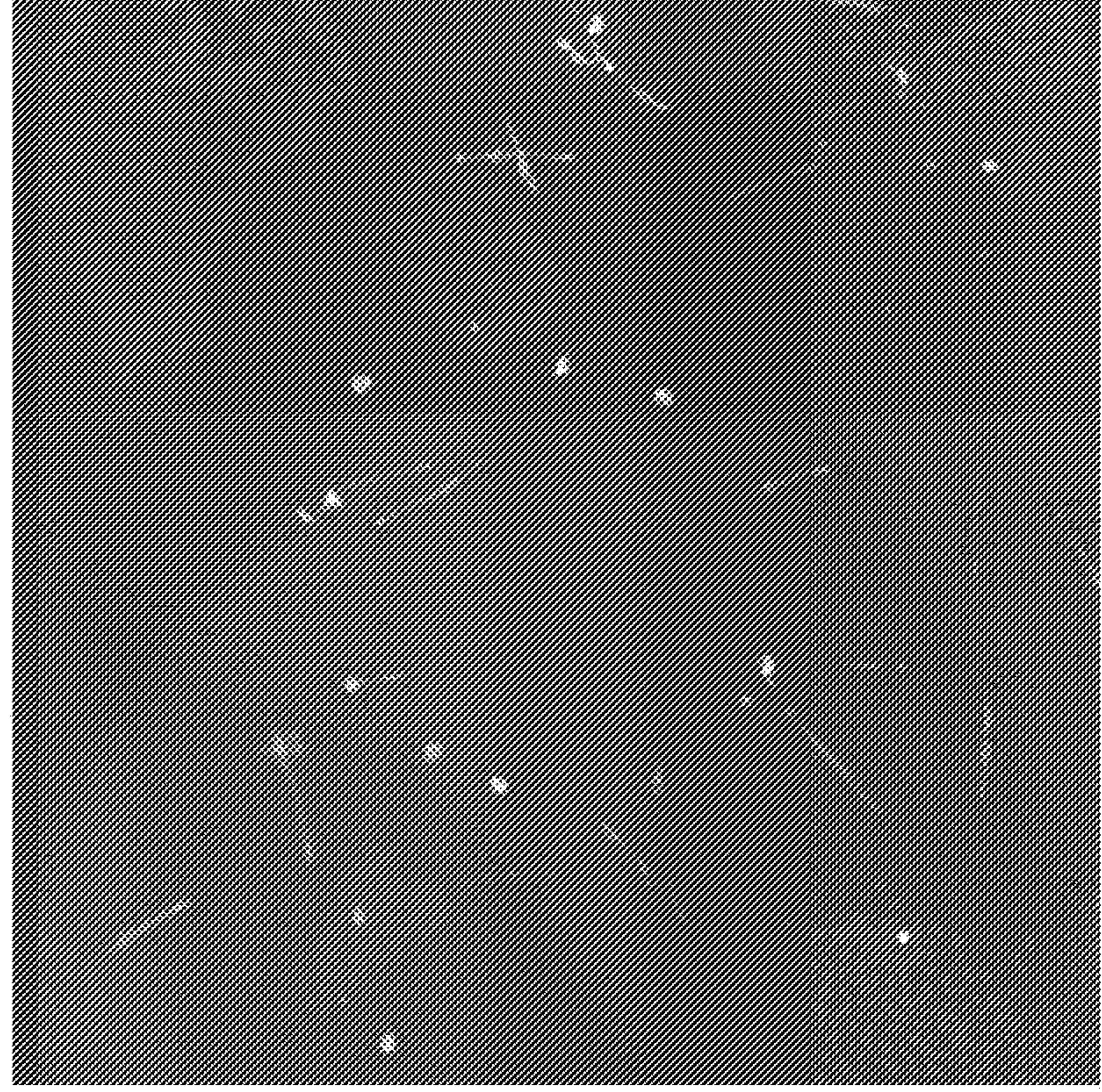
VE-Cadherin staining

**Figure 11. MYOG induces muscle differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

No doxycycline



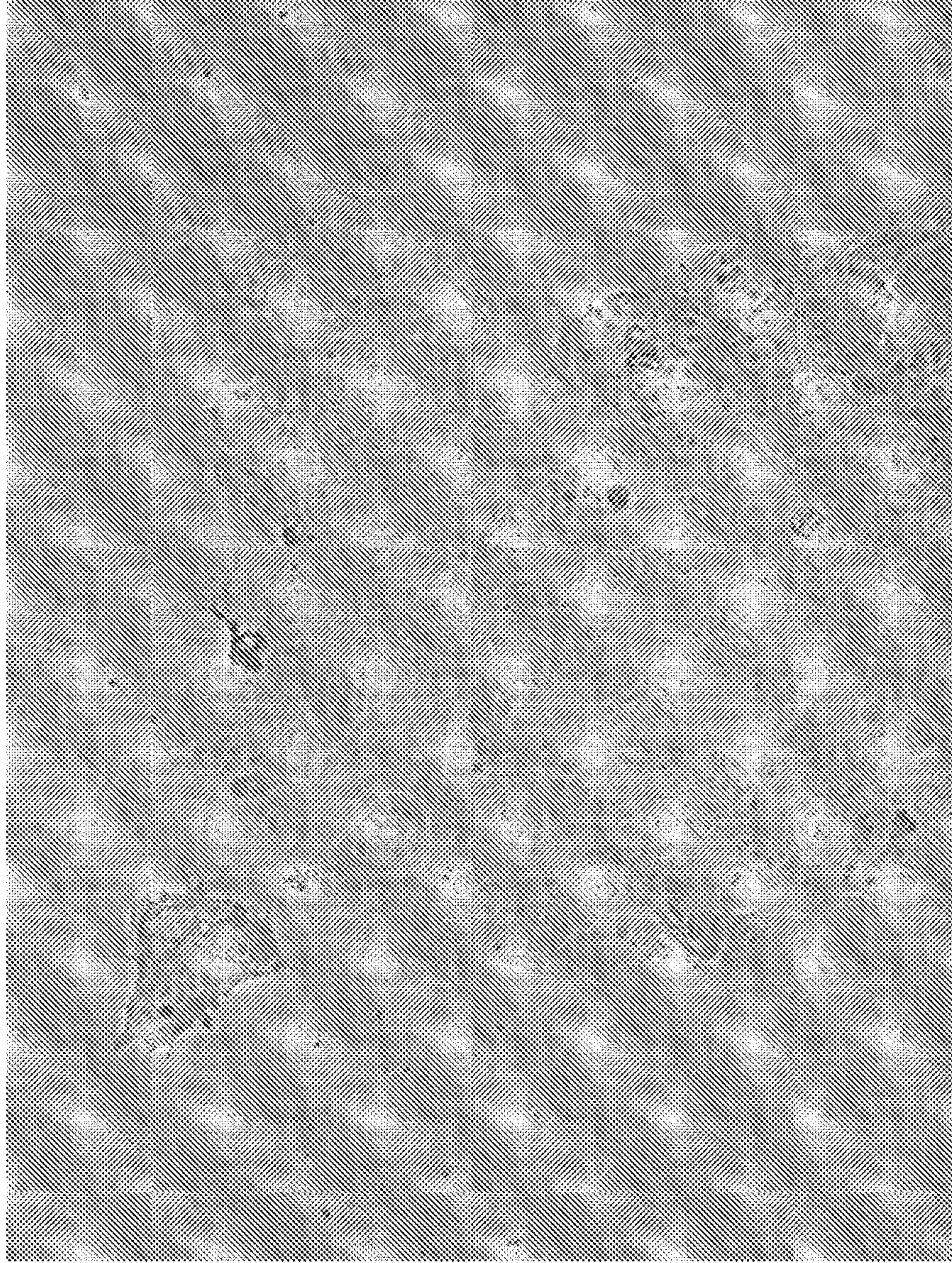
6 days doxycycline



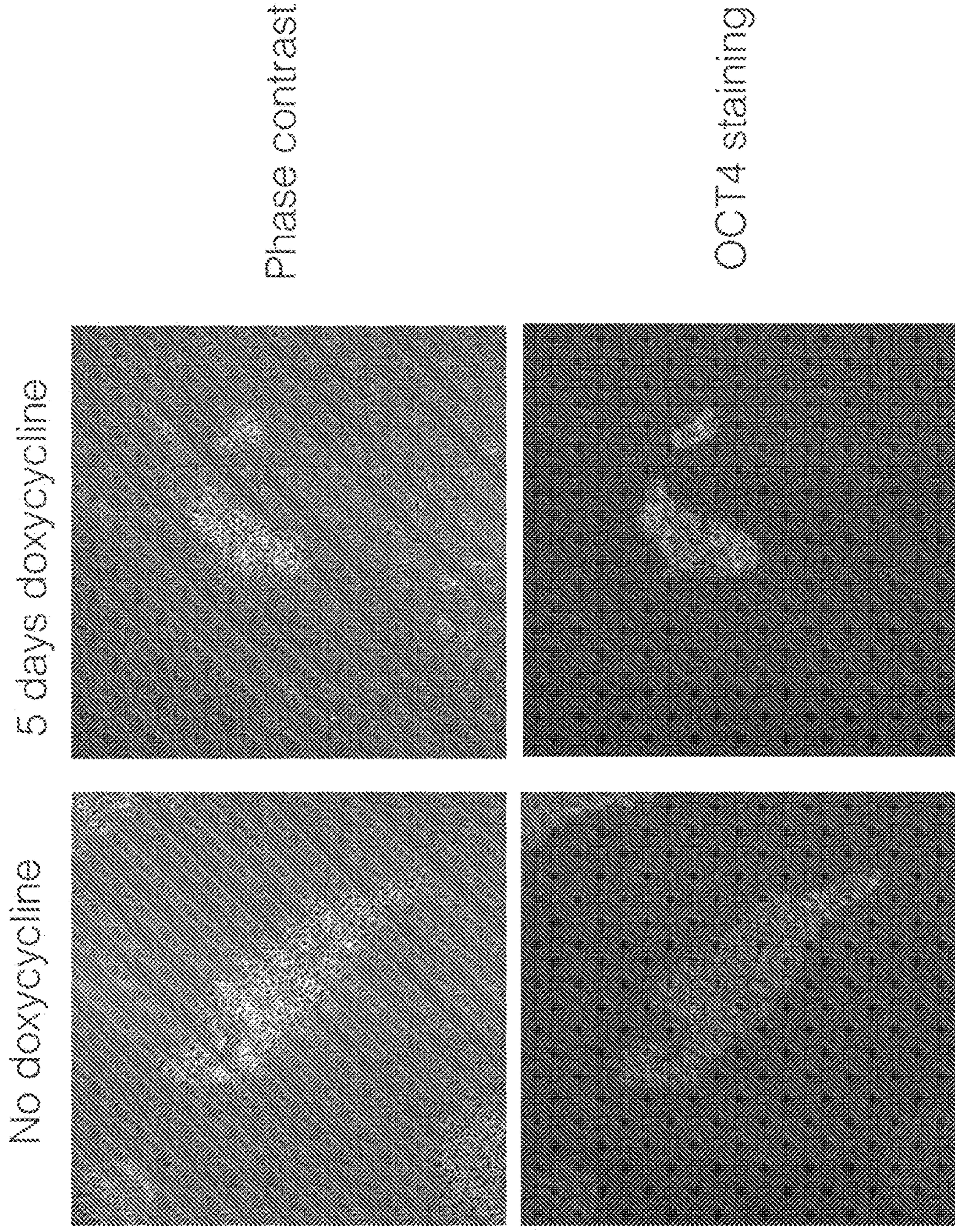
Myosin heavy chain staining

**Figure 12. FOXC1 induces differentiation of hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

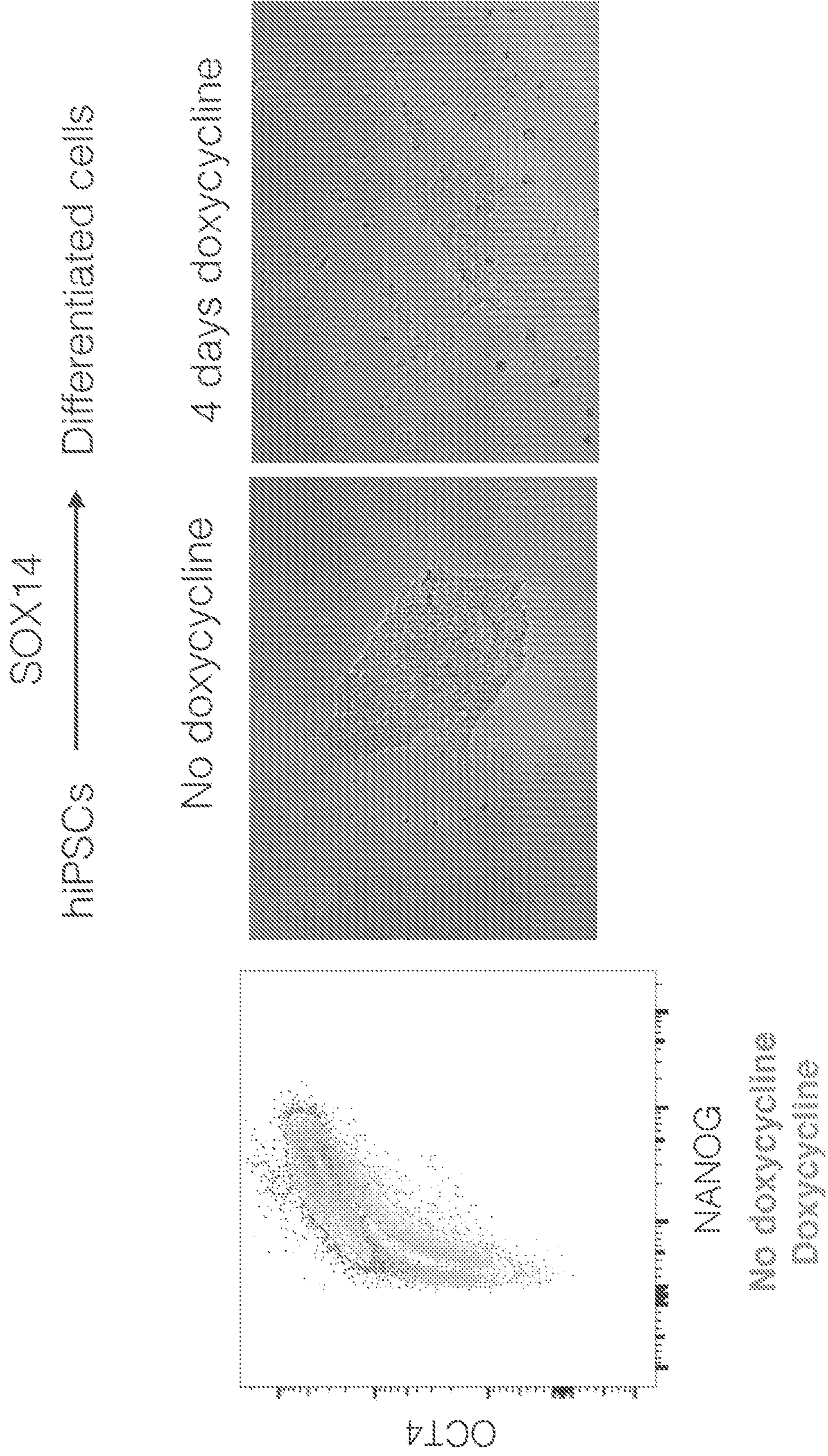
4 days doxycycline



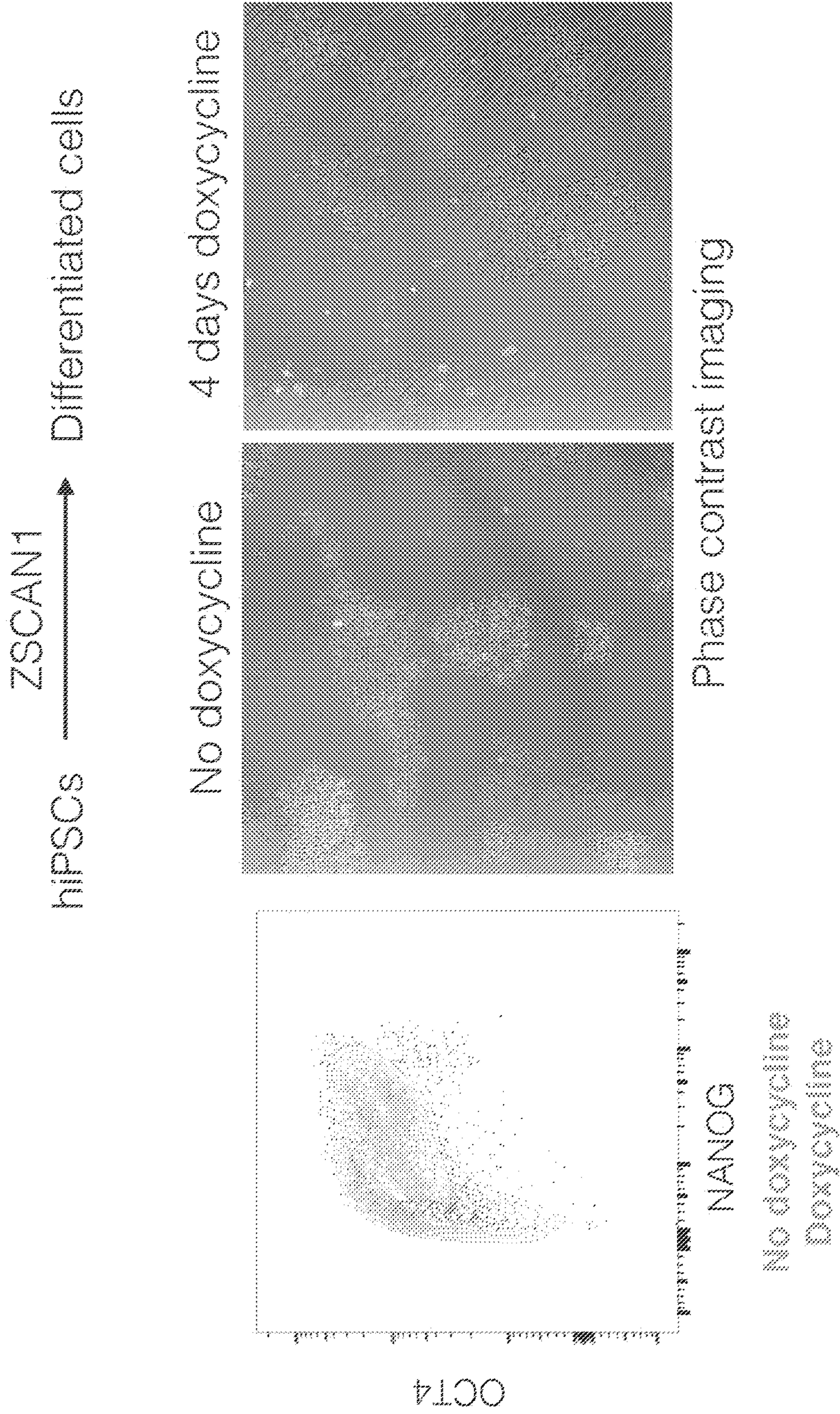
**Figure 13. MITF induces differentiation of hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**



**Figure 14. SOX14 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**



**Figure 15. ZSCAN1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

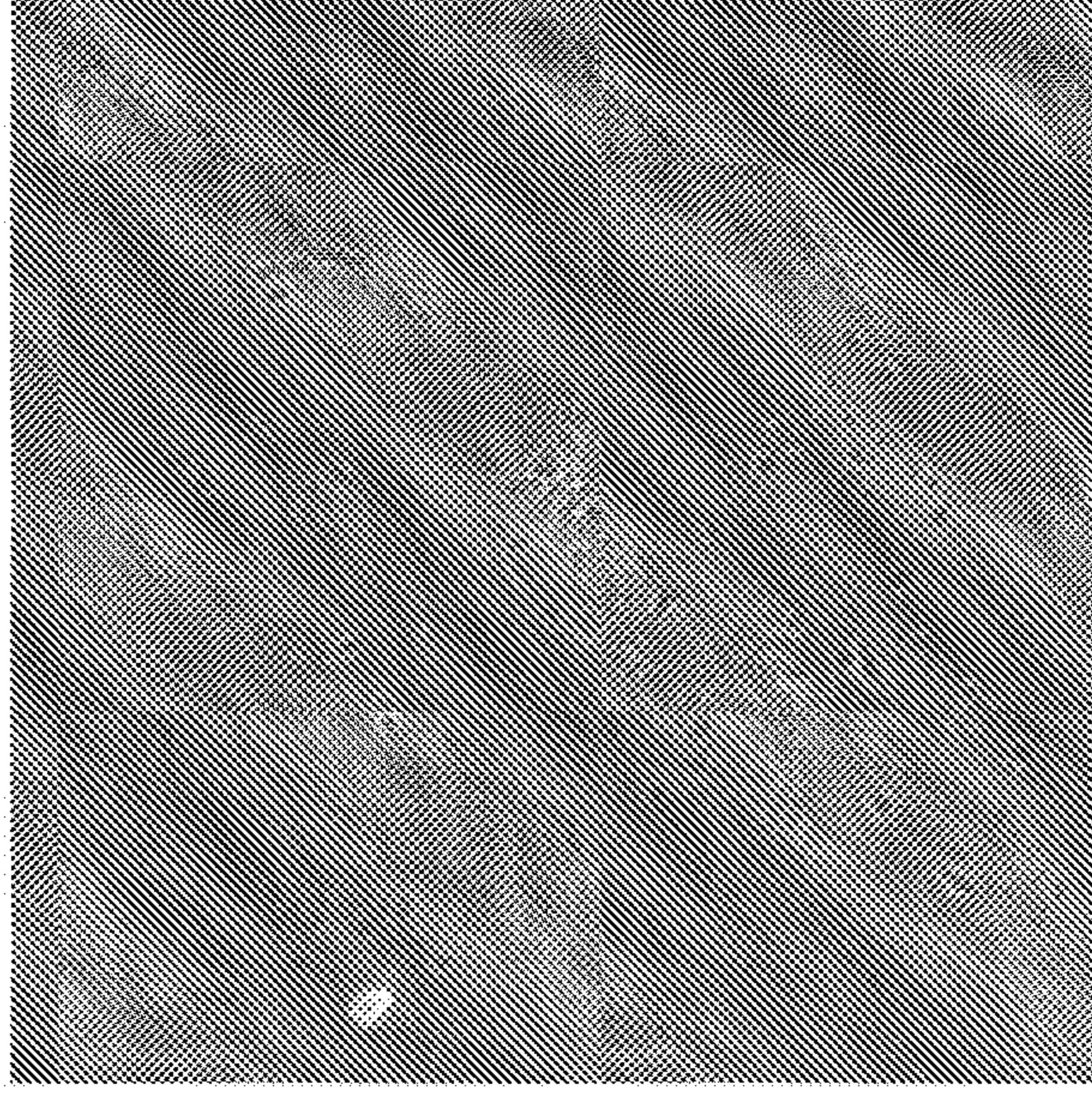
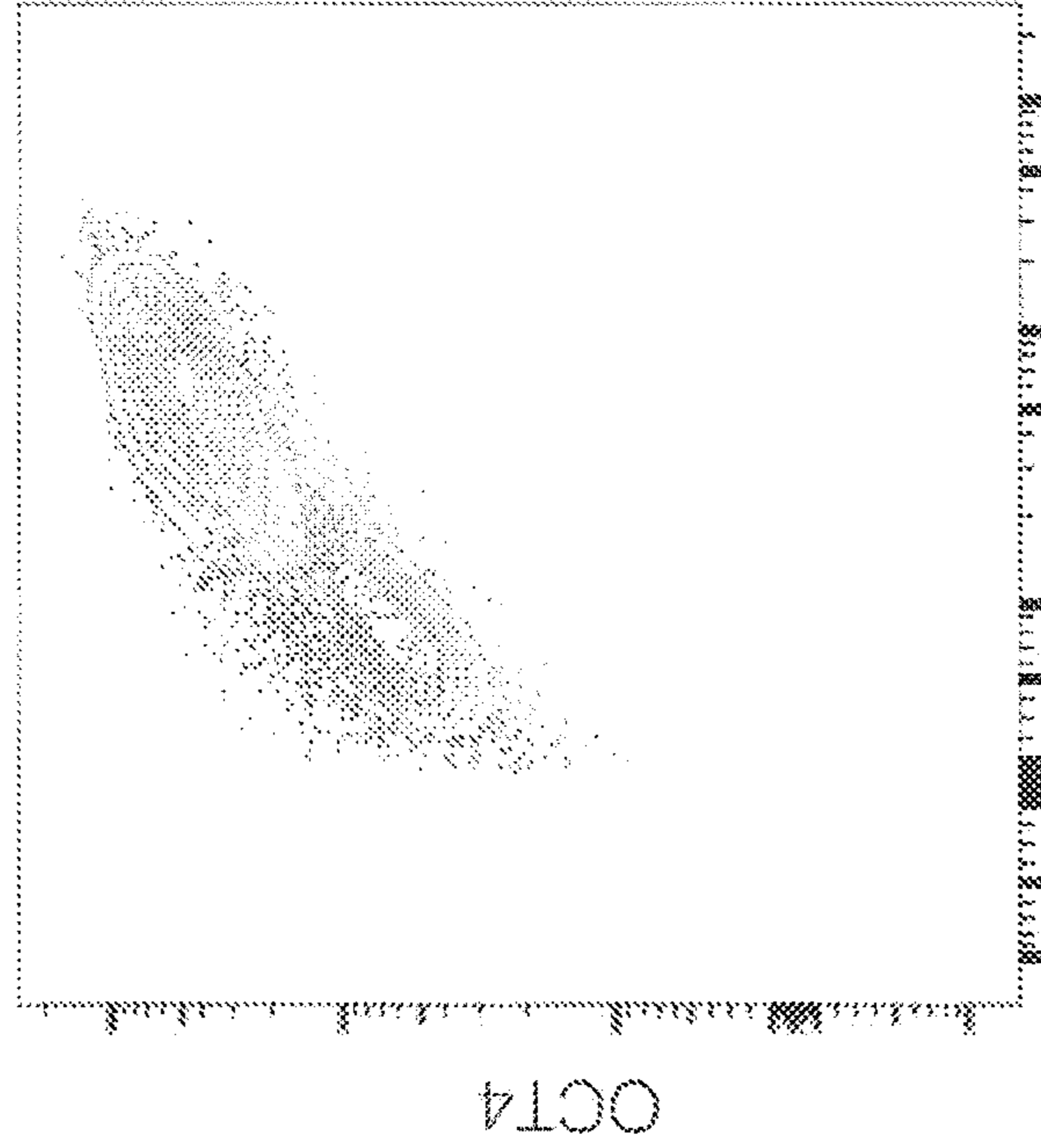




**Figure 16. Wt1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

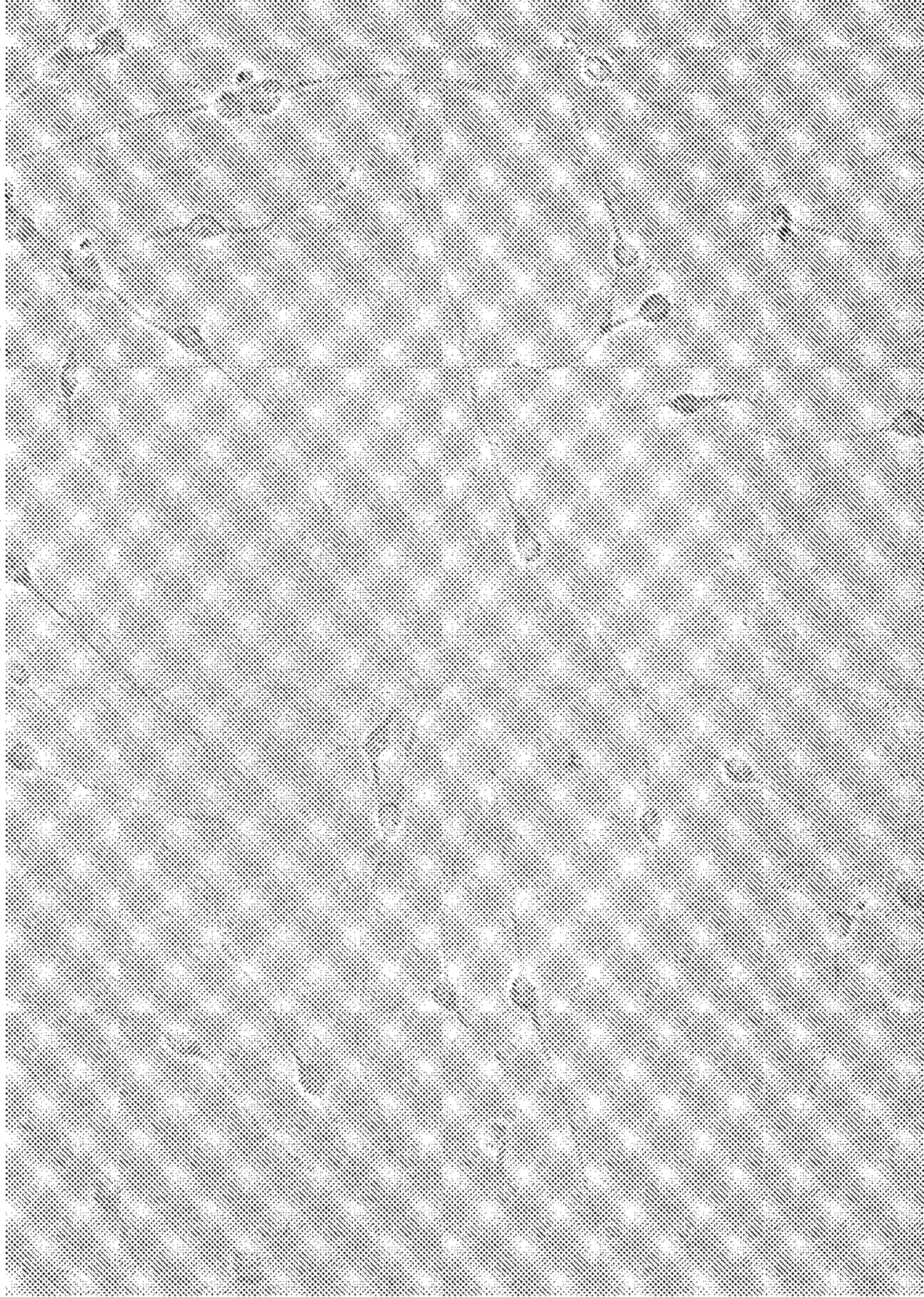
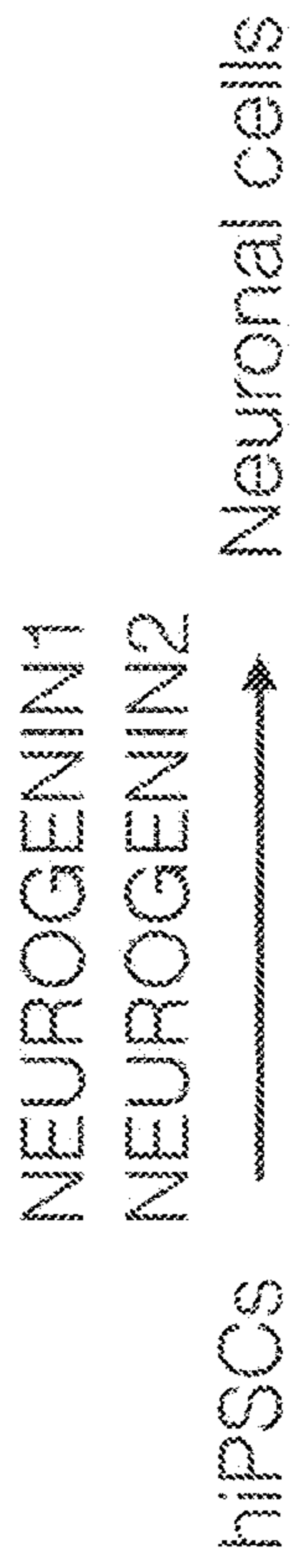


4 days doxycycline



Phase contrast imaging

**Figure 17. NEUROGENIN1 and NEUROGENIN2 induce neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**



**Figure 18. NEUROGENIN1 and NEUROGENIN2 and EMX1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

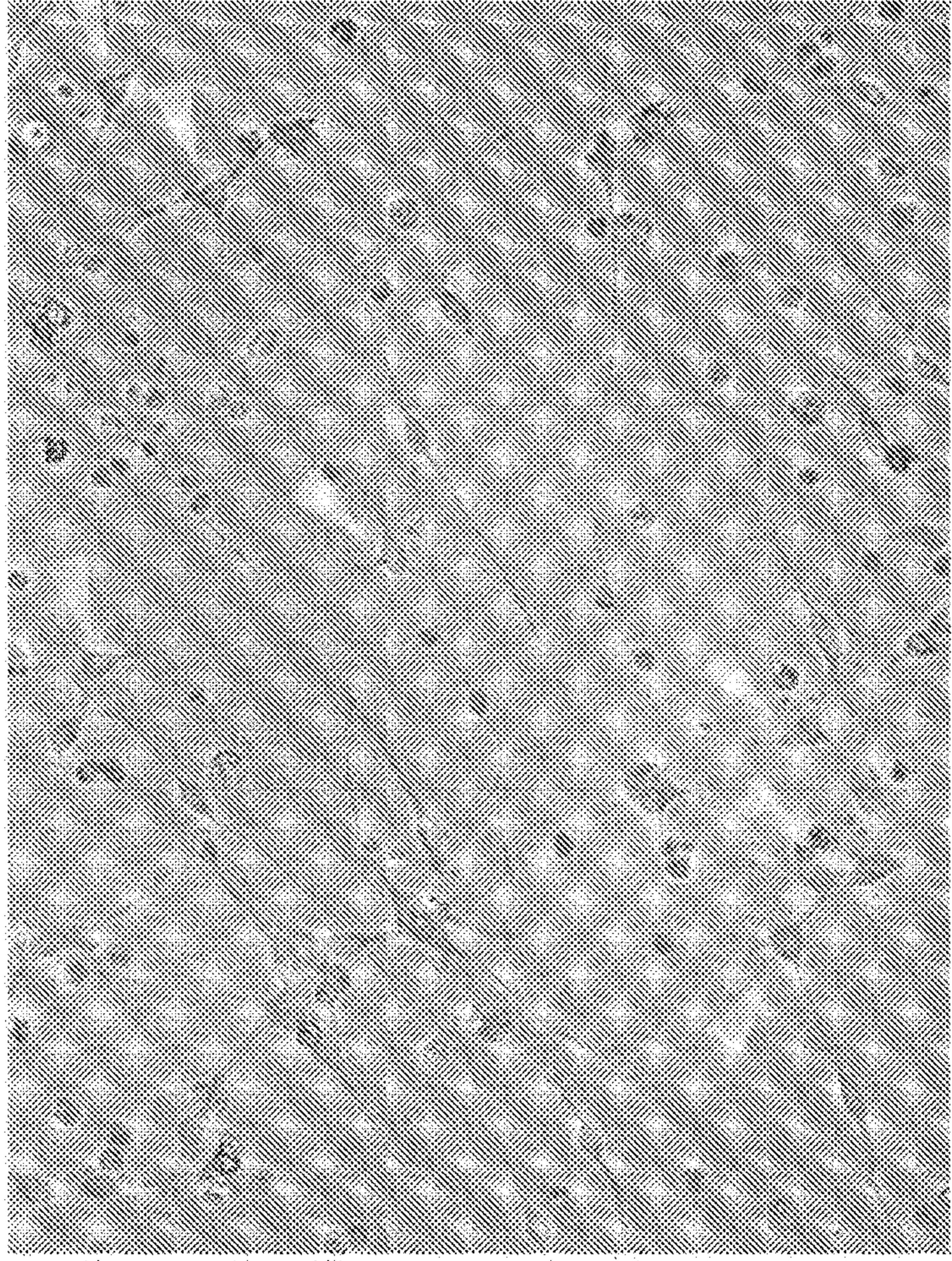
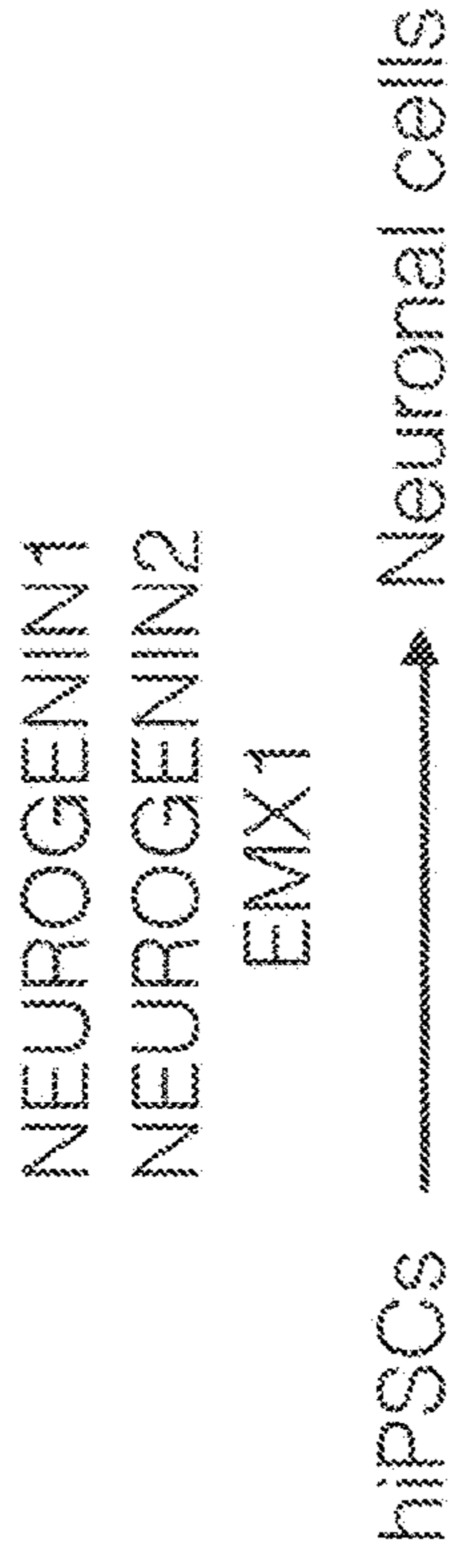


Figure 19. NEUROGENIN1 and NEUROGENIN2 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

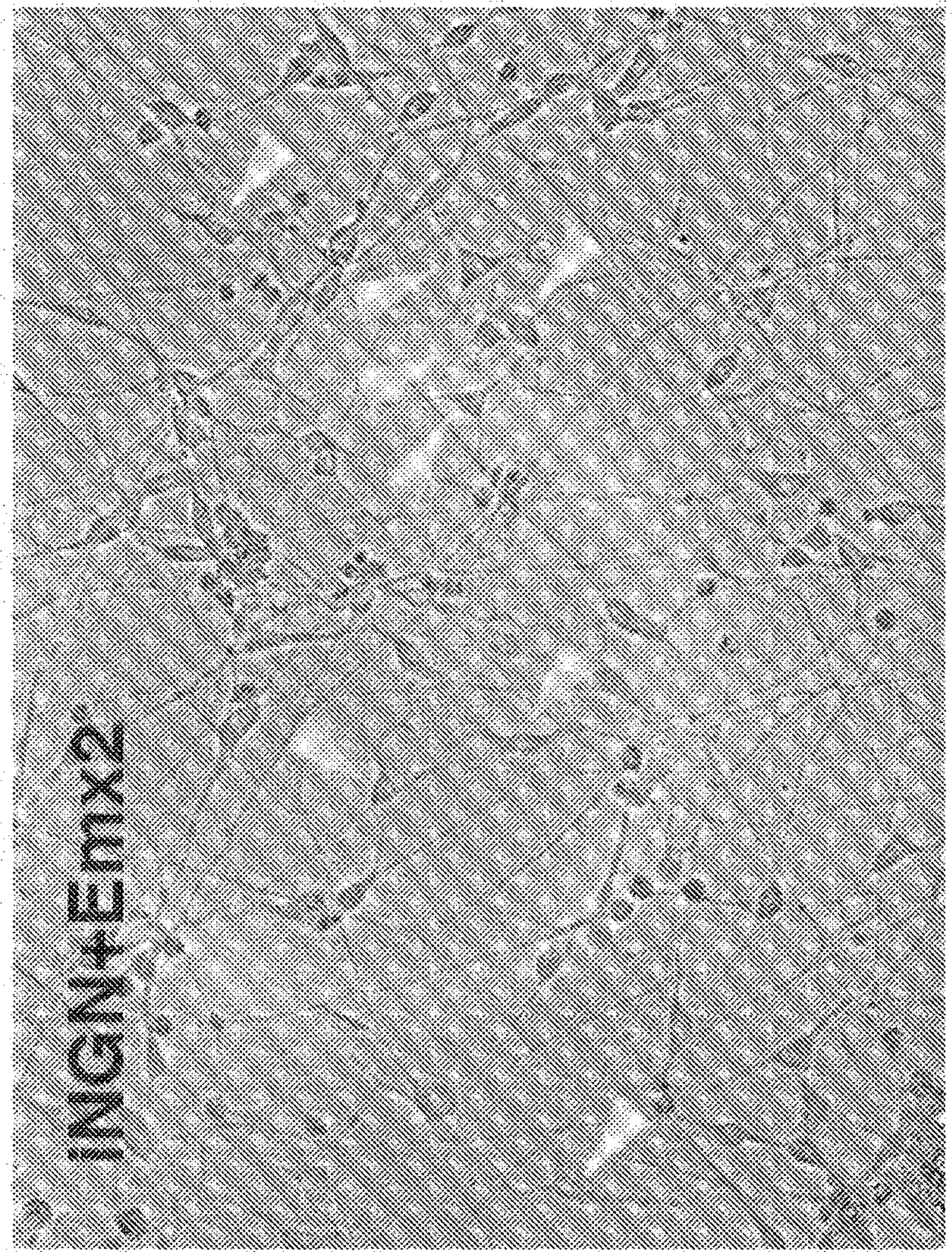
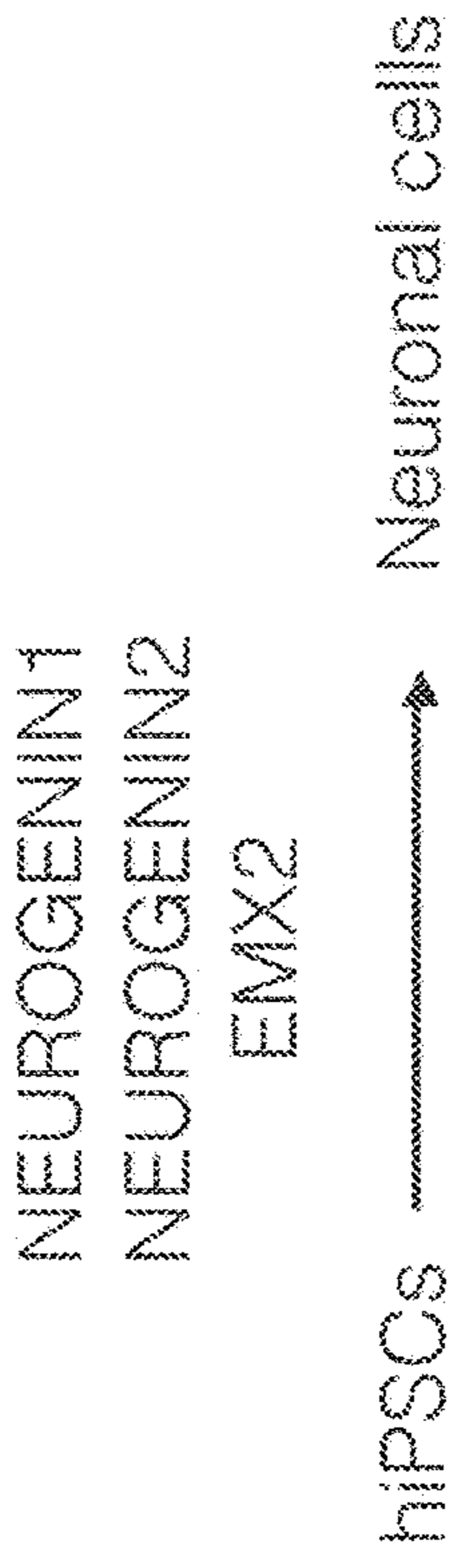
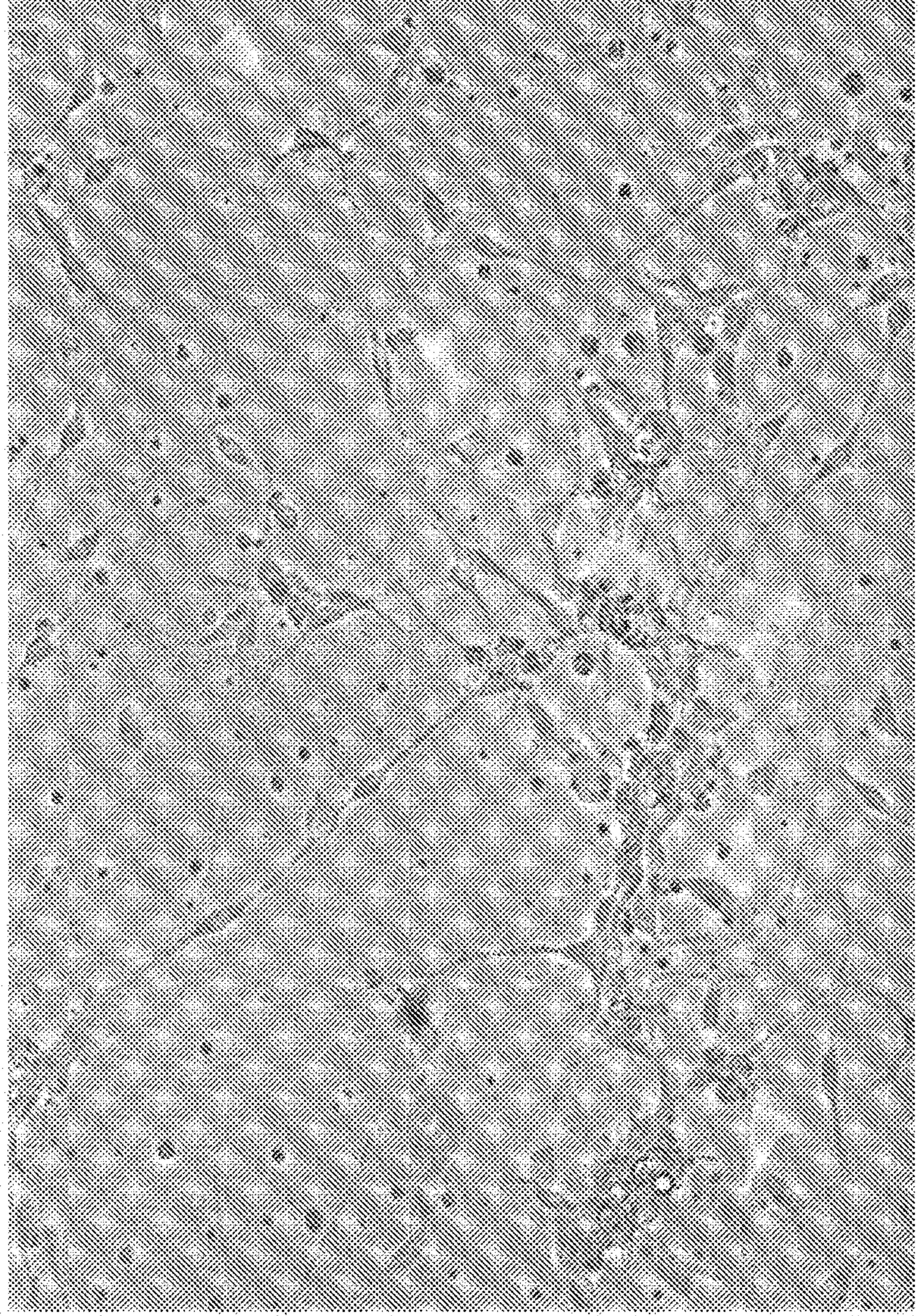


Figure 20. NEUROGENIN1 and NEUROGENIN2 and TBR1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

NEUROGENIN1  
NEUROGENIN2  
TBR1  
hiPSCs → Neuronal cells



**Figure 21. NEUROGENIN1 and NEUROGENIN2 and FOXG1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations**

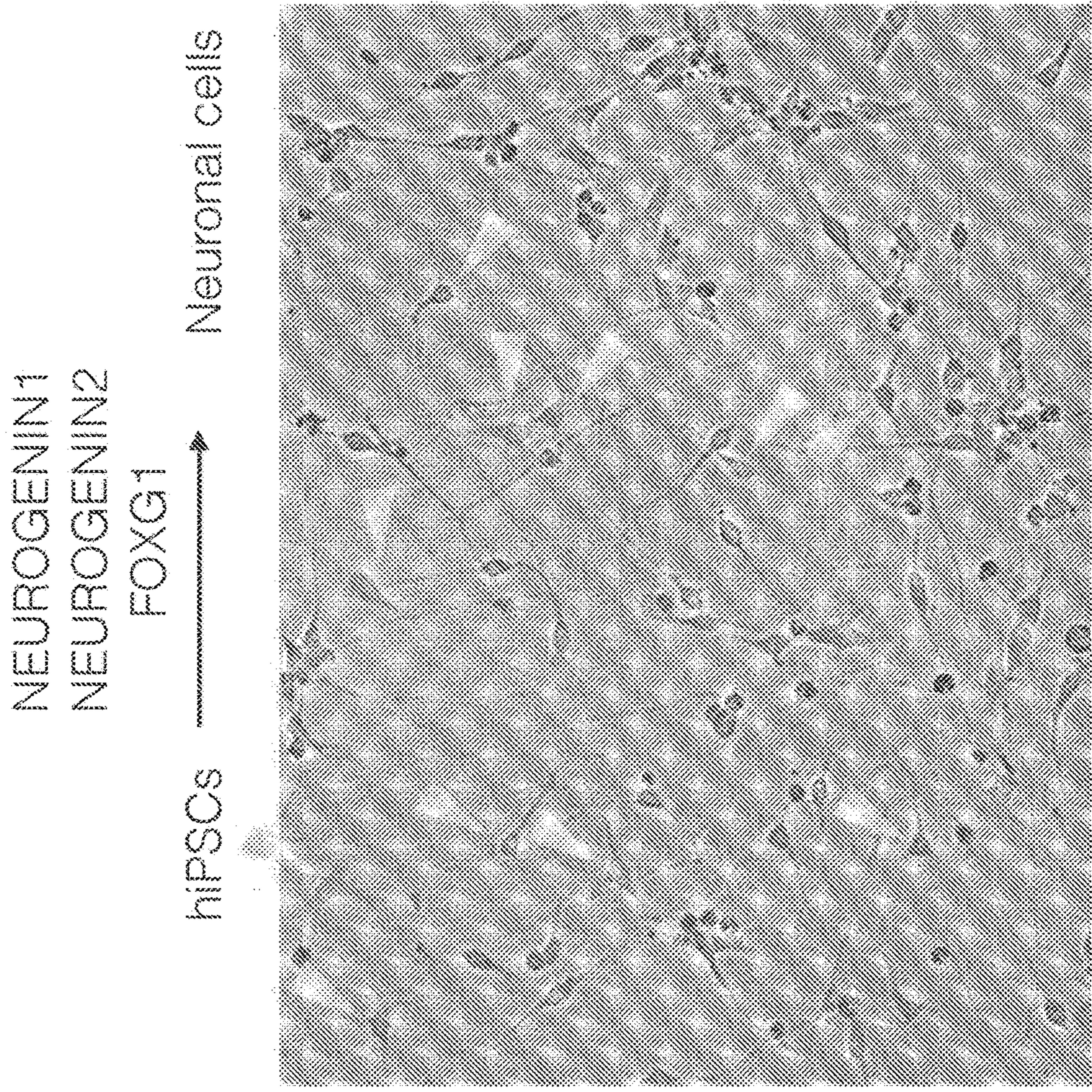


Figure 22. Selected transcription factors (light blue) are lowly or not expressed compared to housekeeping genes (gray) in hiPSCs

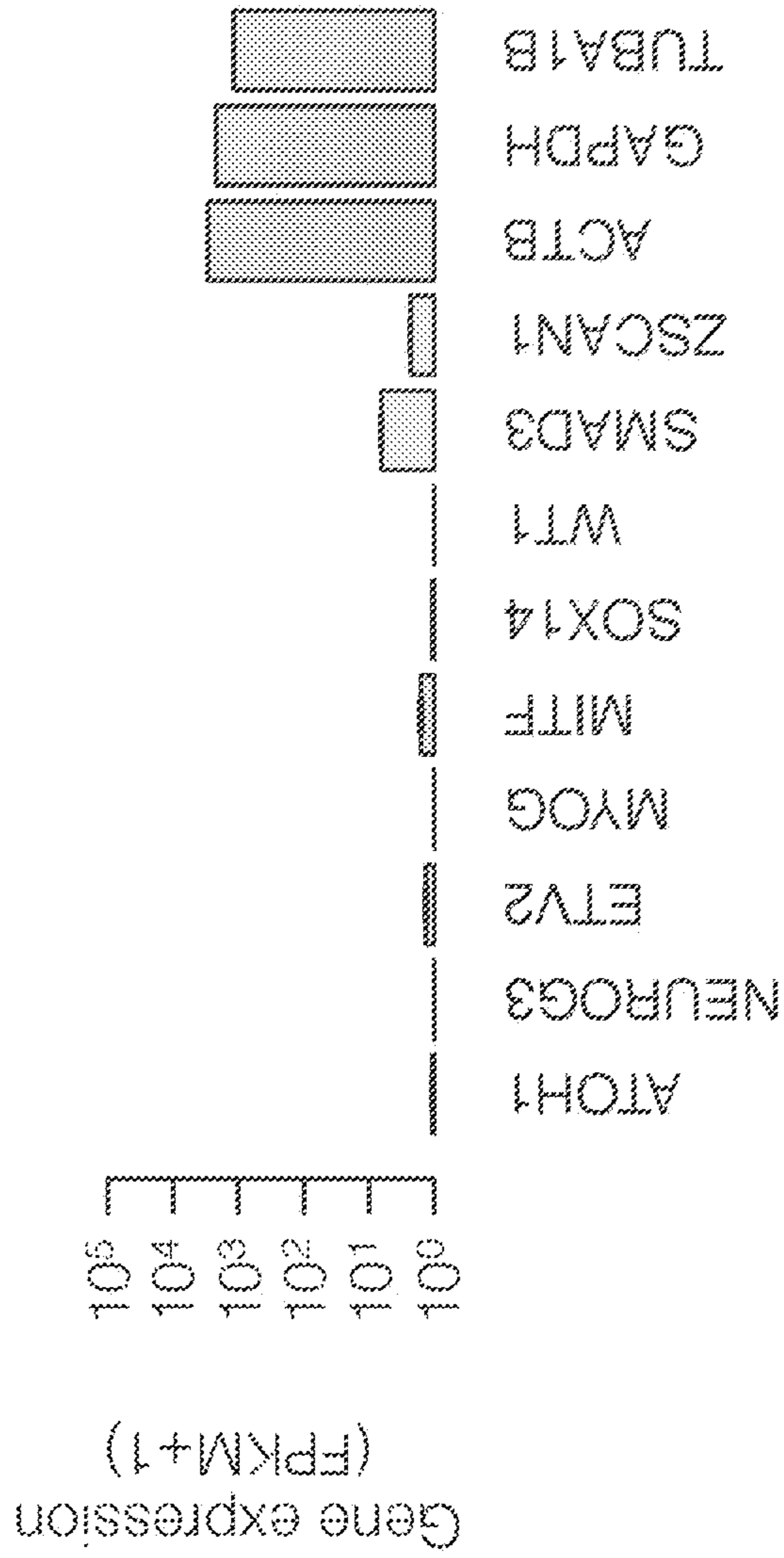
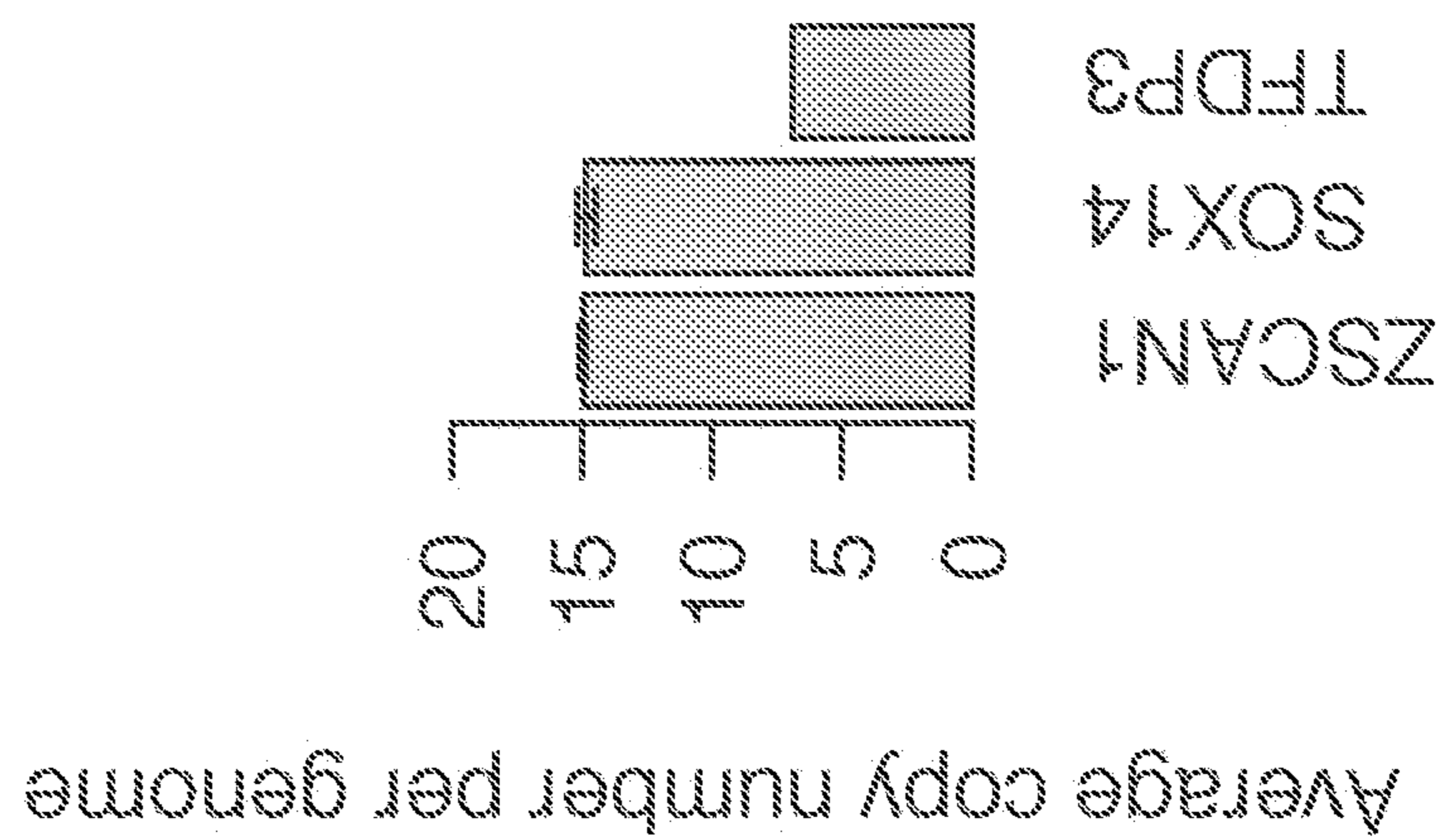


Figure 23. Average copy number of TF ORFs integrated into the hiPSC genome





**TRANSCRIPTION FACTORS  
CONTROLLING DIFFERENTIATION OF  
STEM CELLS**

RELATED APPLICATION DATA

**[0001]** This application claims priority to U.S. Provisional Application No. 62/393,324 filed on Sep. 12, 2016 which is hereby incorporated herein by reference in its entirety for all purposes.

STATEMENT OF GOVERNMENT INTERESTS

**[0002]** This invention was made with government support under 5P50HG005550 and HG008525 awarded by the National Institutes of Health. The government has certain rights in the invention.

TECHNICAL FIELD OF THE INVENTION

**[0003]** This invention is related to the area of cell fates. In particular, it relates to induction of differentiated cells on the one hand and to the maintenance of stem cells on the other hand.

BACKGROUND OF THE INVENTION

**[0004]** The discovery of mouse embryonic stem cells (ESCs), human ESCs and mouse and human induced pluripotent stem cells (hiPSC) has expanded the working modes in biology: unlike studying biological phenomena such as differentiation, gaining and maintaining cellular identity in vivo, we are now theoretically able to mimic some of these processes in a dish. The use of hiPSCs facilitates studying the genesis of human cell types in an ethically approved setting, and enables production of medically relevant cell types for research and medicine. However, exploiting the full differentiation potency of stem cells is only possible with few differentiated cell types. So far, stem cell differentiation protocols are multifaceted and tailored to individual cell types. These protocols often yield highly heterogeneous populations, which may mask the cell type of interest. Whereas the initial triggers that drive stem cells out of pluripotency are known, we know very little about subsequent molecular events that occur during differentiation. For example, initial triggers include the application of differentiation media, the addition of small molecules, application of growth factors or 3D culturing techniques. These stimuli act via cellular signaling cascades and converge on transcription factors (TFs), which alter gene expression by activation and/or repression. Some of these cascades may recapitulate in vivo development, for example retinogenesis in 3D retinal organoids whereas monolayer cultures seem to follow alternative molecular differentiation routes although final cell types can have high similarity with in vivo cell types. Examples of cell identity transitions, both during the course of natural development in vivo and in synthetic in vitro systems, demonstrate the importance of controlling these transcriptional programs. The ability to control cell identity has led to the current preoccupation of the stem cell field, as facile access to primary-like human cells would enable many applications in disease modeling, drug screening and regenerative medicine. Fundamentally, the ability to find and control transitions between cell identities would greatly enhance our understanding of cell identity.

**[0005]** Upon obtaining a cellular identity, its maintenance is a central feature of multicellular organisms. This comprises terminally differentiated cells but also physiological stem cells, which can give rise to specific cell types in the body. To maintain the stem cell pool, transcriptional programs are active in large part by transcription factors that prevent these cells from differentiating. In vitro, maintaining the pluripotency of ESCs or iPSCs is essential to expand and maintain these cells for downstream applications. Labor-intensive and delicate culturing techniques have been developed to maintain pluripotency and to avoid spontaneous differentiation. Advances in the formulation of culture media have significantly improved the robustness of stem cell maintenance, made feeder cell co-cultures dispensable and opened the usage of stem cells to a larger scientific community.

**[0006]** Due to the dependence on transcriptional programs to maintain or convert cellular identities it would be desirable to gain direct control at the transcriptomic level. Detailed study of the transcriptional control of the lac operon was the first example that a certain class of DNA-binding proteins, transcription factors, initiates, enhances or represses transcription. It was shown that one could obtain transcriptional control on cell fate conversion within the same germ layer by ectopic TF expression, which was pioneered by overexpressing MyoD in fibroblasts, which subsequently converted to myoblasts, and C/EBP-alpha to convert B cells into macrophages. Forced TF induction can also convert cells arising from different germ layers, pioneered by the three TFs Brn2, Ascl1 and Myt1l (BAM) that convert fibroblasts into neurons. It has also been shown that certain sets of TF can change the identity of neurons in living animals. Another striking example is the generation of iPSCs by overexpressing Oct3/4, Sox2, Klf4 and c-Myc ("Yamanaka factors") in fibroblasts. TF choices were mostly "biologically inspired;" the ones known from in vivo development were tested in vitro manually or proposed by computational approaches. This biased selection of ectopic TFs led to the successful generation of a handful of cell types. Notably, the failure rate of selected TFs to induce desired cell types from stem cells is relatively high, either not working at all or resulting in unexpected cell fates. Due to experimental and technical differences, for example gene delivery routes, transient or inducible TF expression or differences in starting cell types, we cannot easily troubleshoot these results. Since the forced TF expression in human stem cells can be very efficient, we wondered how one could confer this differentiation route systematically on producing other cell types. Obviously, one needs to standardize the technical aspects of TF delivery, expression, screening and analysis. In contrast to "biologically-inspired" TFs, an unbiased and systematic TF library screen to identify novel TFs that convert stem cells to other cell types or which reinforce pluripotency would be desirable and complementary.

**[0007]** So far, there have not been unbiased, systematic open reading frame (ORF) screens for converting cell identity. More recently, CRISPR-based activator screens have been performed but not for cell conversion. CRISPR-based activators for cell conversion appear to be insufficient to overcome barriers for cell identity conversion. Systematic, RNAi-based screens have revealed important pluripotency factors to understand stem cell biology, but over-expression to maintain pluripotency has not been systematically performed.

**[0008]** Cell types derived from human induced pluripotent stem cells (hiPSC) have high relevance for biomedical research and medicine, but robust, efficient and rapid protocols are lacking for many cell type: could we generate new protocols to generate cell types?

**[0009]** There is a continuing need in the art for methods of changing cell fate and maintaining cell fate so that cell populations available for cellular transplantation and drug screening can better reflect the diverse cell types in the human body.

#### SUMMARY OF THE INVENTION

**[0010]** One aspect of the invention is a method of inducing differentiation of induced pluripotent stem cells. A nucleic acid comprising an open reading frame (ORF) encoding a transcription factor, the transcription factor (TF) protein, or an activator of transcription of the gene encoding the transcription factor, is delivered to the induced pluripotent stem cells. As a consequence, the amount of the transcription factor in the induced pluripotent stem cells is increased, and the induced pluripotent stem cells differentiate to form differentiated cells.

**[0011]** Another aspect of the invention is a method of maintaining pluripotency of induced pluripotent stem cells. A nucleic acid comprising an open reading frame encoding the protein, the protein, or an activator of transcription of an open reading frame encoding the protein is delivered to the induced pluripotent stem cells. The protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells. As a consequence of the delivery, the induced pluripotent stem cells maintain expression of keratin sulfate cell surface antigen, TRA-1-60, which is a marker of stem cell identity.

**[0012]** Another aspect of the invention is an engineered human differentiated cell that comprises one or more nucleic acids comprising an open reading frame. The one or more ORFs encodes a transcription factor selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MTF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1.

**[0013]** Yet another aspect of the invention is an engineered human differentiated cell which comprises a nucleic acid comprising an open reading frame encoding a protein. The protein is selected from the group consisting of ZNF70; ZNF461; TFAP2B; ZNF426; MTF; CDX2; MEOX2; AKNA; NKX2-8; NKX3-2; NKX2-3; ZNF16/HZF1; ETV2; TFPD3; RELA/p65; NEUROG1; ID4; HES7; MXD4; SOX14; FOXP1; E2F2; NEUROG3; ZNF148/ZBP89; GRLF1/ARHGAP35; BOLA2; ZNF616; MAX; ATOH1; PRDM5; LHX5; ZNF273; MAFK; HOXA1; HIF2A/EPAS1; MAFB; E2F3; PRDM7; ZNF44; HMGA1; NRL; BATF3; MYOG; KLF15; LMX1A; HOXB6; DMRTB1; ATF7; SCRT2; ZNF593; HES2; ZSCAN2; MSX2; ID3; SOX12; GLI1; DPRX; SMAD3; ZBED3; CAMTA2; MSC; ASCL3; BARX1; DMRT1; HOXA10; TSC22D3; ZNF837; MXD3; ZNF692; NHLH2; ZNF626; THAP3; SRY; WT1; SHOX; ZNF43; and ZSCAN1.

**[0014]** Still another aspect of the invention is an induced pluripotent stem cell which comprises a nucleic acid comprising an open reading frame encoding a protein. The

protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells. The induced pluripotent stem cell expresses keratin sulfate cell surface antigen, TRA-1-60. The transcription factor may be one or more of the group consisting of AEBP2; AIRE; ARNT2; ARNTL; ATF6; BACH1; BARX2; CASZ1; CLOCK; CREBRF; CTCFL; CXXC1; DLX2; E2F1; EBF1; EGR1; ELF4; ELF5; EMX2; ETS1; ETV1; ETV7; FEZF1; FOXJ3; FOXL2; FOXN2; FOXO1; FOXP3; FOXS1; GATA1; GCM1; HMBOX1; HOXB9; HOXC10; HOXC6; HOXD10; HOXD3; HOXDS; HSF2; HSFY1; IRF5; IRF9; IRX2; KCMF1; KLF1; KLF13; KLF16; KLF7; KLF8; LBX1; LCORL; MAF; MEF2A; MEIS2; MNT; MYBL2; MYF6; MYRF; NANOG; NEUROD1; NFE2L2; NFE2L3; NFIB; NFIX; NKRF; NOTO; NR0B1; NR1B1; NR1B3; NRIC2; NRIC3; NRIF3; NR1I2; NR2B2; NR2C1; NR3A1; NR3A2; NR3C4; NR5A2; OSR1; PAX6; PAX7; PAX9; PHOX2B; PKNOX1; PLAGL1; PREB; PRRX1; RBPJ; RC3H2; RFX5; SATB1; SMAD2; SMAD4; SP1; SP4; SP6; SREBF1; STAT3; TAL2; TBX14; TBX18; TBX21; TBX22; TBX5; TEAD2; TERF1; TERF2; TFAP2A; TFDP2; TFEB; THAP1; THAP10; TP63; TSHZ3; YEATS2; ZBED2; ZBTB12; ZBTB14; ZBTB17; ZBTB21; ZBTB35; ZBTB9; ZEB1; ZFH3; ZFPM2; ZFY; ZFYVE26; ZIC4; ZKSCAN1; ZKSCAN12; ZKSCAN14; ZKSCAN19; ZKSCAN24; ZKSCAN4; ZKSCAN5; ZMAT2; ZNF114; ZNF138; ZNF146; ZNF227; ZNF253; ZNF266; ZNF280A; ZNF280C; ZNF282; ZNF296; ZNF311; ZNF317; ZNF337; ZNF34; ZNF35; ZNF350; ZNF366; ZNF396; ZNF398; ZNF41; ZNF415; ZNF460; ZNF485; ZNF497; ZNF511; ZNF512; ZNF517; ZNF543; ZNF550; ZNF586; ZNF613; ZNF615; ZNF619; ZNF641; ZNF644; ZNF645; ZNF648; ZNF655; ZNF658; ZNF662; ZNF664; ZNF669; ZNF704; ZNF740; ZNF754; ZNF75A; ZNF774; ZNF789; ZNF8; ZNF80; ZNF84; ZSCAN16; ZSCAN22; ZSCAN32; ZXDA; ZXDC; and ZZZ3.

**[0015]** Another aspect of the invention is a method of inducing differentiation of induced pluripotent stem cells. Increased expression is induced of a gene selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MTF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1. The induced pluripotent stem cells thereby differentiate to form differentiated cells.

**[0016]** Yet another aspect of the invention is a method of inducing differentiation of induced pluripotent stem cells. Expression of a gene is induced. The gene is selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MTF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1. The induced pluripotent stem cells thereby differentiate to form differentiated cells.

**[0017]** These and other embodiments which will be apparent to those of skill in the art upon reading the specification provide the art with tools and methods for controlling the cell fate of stem cell populations.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0018]** FIG. 1. Workflow for screening the human TFome (a comprehensive expression library of 1,578 human transcription factor (TF) clones with full coverage of the major TF families) for loss of pluripotency in hiPSCs (human-induced-pluripotent-stem-cell)

**[0019]** FIG. 2. General strategy for high-throughput screening for cell type conversion

**[0020]** FIG. 3. The Human TFome expression library

**[0021]** FIG. 4. Expression vectors for delivery of human TFome

**[0022]** FIG. 5. Hits from high-throughput screening for stem cell differentiation [loss of stem cell identity (TRA-1-60) in human induced pluripotent stem cells]

**[0023]** FIG. 6. Hits are enriched for developmental genes and protein domains

**[0024]** FIG. 7. Selected transcription factors that induce stem cell differentiation

**[0025]** FIG. 8. ATOH1 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0026]** FIG. 9. NEUROG3 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0027]** FIG. 10. ETV2 induces endothelial differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0028]** FIG. 11. MYOG induces muscle differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0029]** FIG. 12. FOXC1 induces differentiation of hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0030]** FIG. 13. MITF induces differentiation of hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0031]** FIG. 14. SOX14 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0032]** FIG. 15. ZSCAN1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0033]** FIG. 16. WT1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0034]** FIG. 17. NEUROGENIN1 and NEUROGENIN2 induce neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0035]** FIG. 18. NEUROGENIN1 and NEUROGENIN2 and EMX1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0036]** FIG. 19. NEUROGENIN1 and NEUROGENIN2 and EMX2 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0037]** FIG. 20. NEUROGENIN1 and NEUROGENIN2 and TBR1 induce cortical neuronal differentiation from

hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0038]** FIG. 21. NEUROGENIN1 and NEUROGENIN2 and FOXG1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

**[0039]** FIG. 22. Selected transcription factors (first nine shown) are expressed at a low level or not at all, compared to housekeeping genes (last three shown) in hiPSCs

**[0040]** FIG. 23. Average copy number of TF ORFs integrated into the hiPSC genome

## DETAILED DESCRIPTION OF THE INVENTION

**[0041]** The inventors have developed techniques for inducing differentiation of stem cells into particular cell lineages, as well as techniques for inducing stem cells to continue to proliferate as stem cells. Rather than using special growth conditions, small molecules, or growth factors, the techniques deploy transcription factors (TFs) to trigger differentiation programs or stem cell renewal programs that prevent differentiation.

**[0042]** As determined, certain transcription factors are able to induce stem cells to differentiate to particular lineages. For example, MITF causes stem cells to form melanocytes. Similarly, CDX2 causes stem cells to form placental cells. MYOG causes stem cells to form smooth muscle cells. An initial indicator of differentiation can be the loss of stem cell specific markers. For example, the keratin sulfate cell surface antigen, TRA-1-60, can be assayed in the cells, as it is lost early in the process of differentiation. Using loss of stem cell markers as an indicator also provides a general means for detecting relevant transcription factors, rather than looking for acquisition of a marker that is newly expressed during differentiation by a particular cell lineage. Loss-of-marker screening may be applied to cell types other than stem cells. For instance, to identify TFs that directly convert fibroblasts into cell types of interest. In some aspects, combination of transcription factors may be used to achieve differentiation to a particular cell lineage. The combination may achieve a cell type or a cell sub-type that is not achieved by either transcription factor alone. Alternatively, the combination may achieve the same cell type as one of the transcription factors alone, but may achieve it more efficiently.

**[0043]** Stem cells may be sorted from differentiated cells or differentiating cells by various means, typically based on differential gene expression. For example, fluorescence activated cell sorting may be used to separate cells on the basis for their expression of TRA-1-60. Alternatively, certain differentiated cells may be sorted from other differentiated cells and from cells on the basis of their expression of a lineage-specific cell surface antigen. Yet another means is by assessing expression at the RNA level, by single cell RNA sequencing without any sorting or pre-selection step. Such techniques are known in the art and may be used as is suitable and convenient for a particular application.

**[0044]** Any means known in the art for increasing the amount of a transcription factor in a stem cell can be used. This may involve delivery of either a nucleic acid comprising an open reading frame encoding the transcription factor, delivery of the transcription factor itself, or delivery of an activator of the transcription factor or its expression. Any technique known in the art for such delivery may be used.

For example, for delivery of a cDNA, a viral or plasmid vector may be used. The open reading frame (encoding any isoform of the TF) may be inducible or repressible for control, to achieve a suitable level of expression. The nucleic acid comprising the open reading frame may be a cDNA, an mRNA, or a synthetic or engineered nucleic acid. Some transcription factors may require a critical amount of expression to effectively induce differentiation, such as the equivalent of at least 5, 10, 15, 20, 25, or 50 copies of the ORF per cell. Other factors may require less than a certain threshold of expression due to possible toxicity at high levels, such as less than 20, 10, or 5 copies per cell. Increased levels of expression may also be achieved by increasing the copy number of an ORF, for example, by using a higher copy number vector or by using a transposon. In some embodiments, nuclease-null or “dead” Cas9 variants may be used to activate the transcription of a desired transcription factor. See, e.g., Chavez et al., *Nature Methods* 13:563-569, 2016. In other embodiments, modified RNAs—RNAs that encode the transcription factor, but use synthetic nucleotides that improve stability and reduce degradation—may be used. In some cases, use of culture media adapted for a particular cell type may increase the expression of the ORF that induces expression of that cell type. Expression of an ORF can be increased from a non-expressed gene, from a gene expressed at a low level, or from a gene expressed at a robust level. Overexpression is expression at level that is higher than the level that is expressed before induction from a gene that is expressed at a low, medium or high level.

**[0045]** An exogenous open reading frame is typically an open reading frame that differs from the similar gene or mRNA in the cell. It may be engineered to have a different control sequence or sequences, such as promoter, operator, enhancer, terminator, etc. It may be engineered to have no introns. It may be engineered to be fused to a second open reading frame to which it is not fused in the human genome.

**[0046]** The differentiated cells that can be produced using the methods described here will have multiple applications. They can be used for regenerative medicine, such as transplanting the cells into a recipient in need of a certain type of cell. They can be used for drug testing, both in cell culture as well as after transplantation. The cells may be used to deliver a product to a part of a body. for example, if they naturally produce or are engineered to produce and secrete the product.

**[0047]** Drug testing in the cells may use substances that are known or unknown to have a certain biological activity. The substances may be elements, compounds or mixtures, whether natural or synthetic. The cells may be used to determine a desirable activity of a potential drug or conversely to determine undesirable effects of a substance or lack of such effects. The contacting of the substance with the cells may be in culture or in a human or animal body. The activity or side effects of the substance may be determined in vitro or in vivo, irrespective of where the contacting occurred.

**[0048]** Changes that are observed in the cells being tested are not limited. The cells can be observed for effects on cell growth, apoptosis, secreted products, expression of particular products. etc. The genome of these cells may be edited to match mutations found in patients with disease. Any type of assay known in the art for such changes may be used, including but not limited to immunological assays, morphological observations, histochemical stains, reverse transcrip-

tion polymerase chain reaction, protein blots, mass spectrometry, hybridization assays, electrophysiology, etc.

**[0049]** The stem cells may be obtained from any source. One particularly useful source is human induced pluripotent stem cells. Mouse induced pluripotent stem cells and mouse embryonic stem cells may also be used, as well as such cells from other animals. The use of human embryonic stem cells may be regulated or ethically undesirable, but these may be used as well.

**[0050]** Differentiated cells may be identified by any property or set of properties that is characteristic or defining of that type of differentiated cell. For example, different cell types have a unique transcriptome. The transcriptome may be used as a means of matching and identifying an unknown cell type to a known cell type. The transcriptome may be used qualitatively or quantitatively. Similarly a proteome may be used a means of identifying an unknown differentiated cell type. Some cell types may be identifiable based on morphology, growth habit, secretion products, enzymatic activity, cellular function, and the like. Any means known in the art for identifying cells may be used.

**[0051]** The above disclosure generally describes the present invention. All references disclosed herein are expressly incorporated by reference. A more complete understanding can be obtained by reference to the following specific examples, which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

#### EXAMPLE 1

**[0052]** We decided to compose a complete TF ORF library by combining factors from existing resources. Absent TFs were obtained by de novo gene synthesis. This “human TFome”, comprises 1,578 canonical human TFs driven by an inducible promoter system within an all-in-one Tet-ON lentiviral vector backbone for stable integration in hiPSCs. By screening the human TFome we found over 70 TFs that induce loss of hiPSC identity, suggesting pervasive potency for TFs to alter cell identity. This resource, the entire library as well as individual TFs, will be publicly available at Addgene (non-profit plasmid repository) and a subset of hiPSC lines in which certain TFs can be induced will be available. We applied the human TFome as single TFs per hiPSCs to identify individual TFs that convert stem cells to other cell identities or reinforce pluripotency. High-throughput fluorescence cell sorting (FACS) in combination with next generation sequencing and subsequent bioinformatic analyses was performed to screen for TFs in two hiPSC lines. We show that expression levels are critical and can be elevated by using transposon-mediated integration of expression cassettes. The application of the human TFome in hiPSCs resulted in the validation of over ten TFs cell fate converting TFs. Furthermore, we discovered over one hundred TFs that reinforce pluripotency, which may help to improve the robustness of stem cell cultures.

#### EXAMPLE 2

##### Generating and Applying the Human TFome

**[0053]** Systematic and comprehensive TF-wide induction screening in human stem cells requires the availability of a TF-expression library. Notably, only partial human libraries, for example TFs included in the ORFome, were accessible.

Therefore, we decided to assemble TFs from available resources or by de novo gene synthesis to compile the “human TFome”, an expression library of 1,578 TFs representing all canonical human transcription factors (Vaquerizas et al. 2009) and further curated.

**[0054]** The pLIX403 lentiviral vector was chosen because it allows for genomic transgene integration, doxycycline-inducible TF expression from a Tet-On system and puromycin selection of transduced cells. The individual TFs in shuttling vectors (pENTR) were cloned into pLIX403 by pooled gateway cloning. To discriminate the ectopic TFs from intrinsic expression, the TFs were marked by a V5 epitope tag translated on the backbone downstream of the Gateway recombination site on the C-terminus of the TF. The pLIX403 vector comprises a second ubiquitous promoter cassette driving the rTA3 gene needed to activate the TetOn promoter in the presence of the small molecule doxycycline and bicistronically a puromycin selection marker. About 98% of the TFs were detectable by next-generation sequencing after subcloning in the DNA plasmid library, which subsequently was used to produce lentiviral particles.

**[0055]** We transduced the PGP1 from the Personal Genome Project and ATCC DYS0100 human iPSC lines with the human TFome at a low multiplicity of infection (MOI=0.1), such that cells would receive a single lentiviral integration at most. In total, we transfected both cell lines each with a complete human TFome pool as well as with two subpools that we have six independent transductions. To obtain sufficient coverage of the library, we ensured that on average each TF was present in at least one hundred cells after lentiviral transduction. Genes that confer resistance to the antibiotics bleomycin, blasticidin, and hygromycin, which should not induce differentiation, were also transduced as negative controls. Transduced cells were then selected for TF-integration using puromycin, expanded and doxycycline was added for four days for continuous TF induction. FACS separated the stained population into the differentiated TRA-1-60<sup>low</sup> and the pluripotent TRA-1-60<sup>high</sup> population. For each of these populations, the integrated genes were amplified using universal primer PCR, and sequenced.

### EXAMPLE 3

#### Transcription Factors-Wide Screening for Alterations in Pluripotency in Human Induced Pluripotent Stem Cells

**[0056]** To identify TFs that induce stem cells into any differentiated cell, we devised a strategy to screen for individual TFs that potently cause loss of pluripotency rather than enriching for cells with any specific cell type marker. We used a fluorescence-activated cell sorting (FACS) approach, which enables multiplexed assessment of the human TFome followed by TF identification by using next-generation sequencing. To measure the loss of pluripotency, we stained for the keratin sulfate cell surface antigen TRA-1-60, which is rapidly lost upon exit from pluripotency. Importantly, because we aimed to identify potent inducers of differentiation, we performed the screen in the standard mTeSR1 stem cell media.

**[0057]** We first aimed to comprehensively identify single TFs that would induce loss of pluripotency. To score each TF for its differentiation potential, we computed a ratio of each

TF by dividing the number of normalized reads sequenced in the differentiated TRA-1-60<sup>low</sup> population versus the pluripotent TRA-1-60<sup>high</sup> population. We expect a TF that induces differentiation to be highly enriched in the TRA-1-60<sup>low</sup> population compared to the pluripotent TRA-1-60<sup>high</sup> population, and hence have a high ratio. Conversely a TF that reduces spontaneous differentiation, for instance by maintaining pluripotency, would have fewer reads in the differentiated population compared to the pluripotent population. We set a threshold for a TF having no effect based on the score of the truncated fluorescent proteins. TFs that increase cell proliferation without affecting pluripotency would be expected to have differentiation scores similar to these truncated fluorescent proteins.

**[0058]** Based on this differentiation scoring metric, we identified over 70 TFs that induce loss of pluripotency (“differentiation-inducing TFs”) and over 100 TFs that reinforce pluripotency (“pluripotency-reinforcing TFs”), as compared to the truncated fluorescent protein controls. To assess the quality of this screen, we conducted gene set enrichment analysis (GSEA) on the differentiation-inducing TFs. We discovered that differentiation-inducing TFs are enriched for developmental processes and system development, which is consistent with their ability to induce loss of pluripotency in our screen. In terms of protein domains, basic helix-loop-helix DNA-binding domains, which are often present in TFs involved in development were enriched, whereas, zinc fingers and Küppel-associated box (KRAB) domains were depleted. Overall, these results suggest that our screen in human stem cells recovers developmentally important TFs that have been identified in model organisms in vivo.

### EXAMPLE 4

#### Transcription Factors Induce Rapid and Efficient Differentiation When Expressed at High Levels

**[0059]** To craft genetic recipes that rapidly and efficiently induce hiPSCs into cell types, we engineered inducible single-TF expressing cell lines that could produce cell types of interest on demand. We originally aimed to mimic closely the conditions of the screen, namely single-copy lentiviral integration. Our initial efforts successfully validated the screen, however the differentiation efficiencies were weak (~10%), similar to CRISPR-Cas9 activator-based differentiation experiments. To overcome this challenge while keeping with our goal for simple differentiation protocols without additional growth factors, mechanical steps or selections, we wondered whether low TF expression levels of the integrated lentiviral vector was the bottleneck of highly efficient differentiation. We surmised that certain crucial target genes may be occluded with a low probability of TF binding and activation of transcription; thus a higher expression would in theory increase the probability of a successful binding and hence transcriptional event that may activate positive feedback loops that induce exit from pluripotency.

**[0060]** To express high levels of TF, we first tested transduction with high titer lentiviral particles, which resulted in massive cell death. To improve the transduction efficiency and the viability of hiPSCs, we constructed a PiggyBac transposon-based vector with similar doxycycline-inducible TF expression and the ability to use puromycin to select for transduced cells. Critically, this transposon system allows for facile high-copy integration of TFs; in our hands, we

average ~15 integrated copies per genome, as assessed by digital droplet PCR. Due to this integration efficiency, it is challenging to set copy numbers to a single TF per cell. Therefore, we decided not to repeat the screen with the lentiviral system but selected TFs that had a high differentiation score for subcloning into the PiggyBac vectors.

**[0061]** Using this high expression system, we assessed differentiation efficiency. By bright-field microscopy, we observed a rapid loss of stem cell morphology, specifically migration away from colonies and adoption of distinct cell morphologies. We quantified loss of stem cell identity by intracellular flow cytometry for NANOG, OCT4 and SOX2. Uninduced stem cells had >90% NANOG<sup>+</sup> OCT4<sup>+</sup> SOX2<sup>+</sup> cells, whereas doxycycline-induced cells had <10% NANOG<sup>+</sup> OCT4<sup>+</sup> SOX2<sup>+</sup> cells.

#### EXAMPLE 5

##### Systematic Classification of TF-Induced Lineages

**[0062]** Next, we needed an approach of identifying a priori the cell lineage being generated—we needed a method analogous to BLAST, but for tissue profiles. Several studies have been published that generally compare gene expression profiles to infer common drug targets, disease mutation effects, etc; however, the broad range of genes did not allow for rigorous and unambiguous identification of cell lineage against expression profiles catalogued in the Gene Expression Omnibus. A BLAST-like approach also requires an extensive reference panel to compare to; however the mechanistic gene regulatory network-based algorithm CellNet requires many samples per tissue, which is not currently available. We systematically use thousands of RNA-seq datasets from many tissues as training data. We adapted KeyGenes, a machine-learning-based algorithm, to systematically BLAST our transcriptomes against high-quality tissue expression datasets.

**[0063]** The TFs that induce loss of stem cell identity may be caused by conversion into differentiated cell types, or general loss of cell identity without a specific identity. To systematically determine what lineage these TF may be inducing, we first needed a systematic approach to classify cell types. We curated a set of large-scale human tissue expression profile studies to train a machine learning classifier for cell types. This curation comprises RNA-seq samples representing tissues from the GTEx study, Human Protein Expression Atlas and Illumina Human Body Map. We then applied this as training data for KeyGenes, LASSO regression-based classifier for cell type identification.

**[0064]** To determine what cell lineages are being generated by these TFs, we performed RNA-seq on each cell population and used them as query for the expanded KeyGenes classifier.

**[0065]** Interestingly, expression of these TFs in adult human tissue did not predict which lineages were produced in hiPSCs. For instance, ETV2 induces endothelial differentiation but is most highly expressed in the testis, ATOH1 activated a neuronal cell identity but is highly expressed in the colon and small intestine, and CDX2 induced a placental identity, but is also highly expressed in the colon and small intestine. Thus observational studies alone do not appear to predict which TFs can induce which lineages, and suggest that synthetic methods of converting cell type do not necessarily recapitulate in vivo development.

**[0066]** Expression of specific markers for these cell lineages was present from the RNA-seq data). To enhance differentiation into these lineages, we induced TF expression in the presence of standard culturing conditions for those lineages, and assess conversion efficiency for lineage-specific markers by flow cytometry. Morphologies and lineage-specific marker expression were tested by immunohistochemical stainings. Together, these data indicate that TF-derived cells can be potently and rapidly generated.

#### EXAMPLE 6

##### Experimental Methods

##### Annotation and Manual Curation of the TFome

**[0067]** Canonical human transcription factors were previously annotated by Vaquerizas et al. 2009) based on a computation search for genes that bind DNA in a sequence-specific manner, but are not enzymatic and do not form part of the core initiation complex, resulting in 1,591 TFs (classified as “a” and “b”, which have experimental evidence for regulatory function, and “other” as probable TFs with undefined DNA-binding domains; class “x” was excluded as having promiscuous DNA-binding domains). We further included TFs that were predicted but did not have experimental evidence (class “c”). The following major TF families were filled in according to HUGO Gene Nomenclature Committee (HGNC): zinc finger (includes C2H2-containing domains), homeodomain (includes LIM, POU, TALE, HOXL, NKL, PRD sub-families), basic helix-loop-helix and forkhead. Pseudogenes as annotated by HGNC were removed. Duplicated and unmapped genes were removed, and all genes were converted to approved gene names using the HGNC multi-symbol checker. The final target set of TFs in the human TFome contains 1,578 genes.

##### Assembly and Quality Control of the Human TFome

**[0068]** All TFs are cloned in pDONR-series standardized Gateway-compatible vector, building on Yang et al. 2011, Jolma et al., transOMIC technologies cDNAs, ORFs (transOMIC Inc), DNA Repository at Arizona State University, and codon-optimized synthesis by Gen 9.

##### Pooled Cloning of the Human TFome into pLIX403 Viral Vector

**[0069]** To perform pooled LR cloning into the pLIX403 viral expression vector (Addgene 41395), each 96-well plate of DNA was combined into its own sub-pool and quantified using Qubit dsDNA Broad Range Assay Kit (Invitrogen Q32853). 75 ng of pENTR-TFome subpool was used for a LR Clonase II (Invitrogen 11791100) reaction overnight and digested with Proteinase K. 1  $\mu$ L of the reaction was transformed into One Shot Stbl3 chemically competent cells (Invitrogen C737303), plated onto a 50 cm<sup>2</sup> LB Agar plate with 100  $\mu$ g/mL Carbenicillin and incubated overnight at 30° C. Serial dilutions were performed to estimate the number colonies per plate. Approximately 10,000 to 20,000 colonies grew per plate, representing 100 to 200-fold coverage per 96-member sub-pool.

**[0070]** Colonies from each sub-pool plate were scraped and DNA was extracted using QIAGEN Plasmid Plus Midi Kit (QIAGEN 12943) to generate pLIX403-TFome sub-

pools. To cloning efficiency and library coverage, pENTR and pLIX403 subpools were pooled to generate pENTR-TFome and pLIX403-TFome. 5  $\mu\text{g}$  of each pool was sheared to 200 bp on a Covaris S2, and 1  $\mu\text{g}$  was used for library preparation using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB E7370L).

#### Lentiviral Production and Transduction

**[0071]** Lentiviruses were produced as before using 45  $\mu\text{g}$  DNA for two 15 cm dishes. hiPSCs were transduced with TFome lentiviruses at a multiplicity of infection (MOI)=0.1.

#### Flow Cytometry Analysis and Fluorescence Activated Cell Sorting (FACS)

**[0072]** Cells were trypsinized, washed and resuspended in FACS buffer (PBS with 10% FBS). For surface antigens, live cells were stained with fluorophore-conjugated antibodies and viability dye CellTrace Calcein Blue, AM (Life Technologies, C34853) at  $1 \times 10^7$  cells/mL for 30 minutes on ice in the dark. For intracellular staining, cells were fixed using BD Cytfix fixation buffer (BD Biosciences, 554655) at  $1 \times 10^7$  cells/mL for 20 minutes, washed with BD Perm/Wash buffer (BD Biosciences, 554723), and permeabilized in BD Perm/Wash buffer for 10 minutes, then stained with antibodies and DAPI in the dark for 30 minutes. Stained cells were washed twice with FACS buffer, filtered into a strainer-capped tube (Falcon, 352235) and run on a BD LSRFortessa. Compensation for spectral overlap was determined by staining AbC Total Antibody Compensation Beads (Life Technologies, A10497) with single fluorophore-conjugated antibodies.

#### Transcriptome Library Preparation and Sequencing

**[0073]** TRIzol (Life Technologies, 15596-018) was added directly to cells and incubated for 3 minutes and used for RNA extraction using Direct-zol RNA MiniPrep (Zymo Research, R2050). RNA was quantified by Qubit RNA HS Kit (Molecular Probes, Q32852). 1  $\mu\text{g}$  was used for Poly(A) isolation using Poly(T) beads (Bioo Scientific, 512979) and used for RNA-seq library preparation with unique molecular identifiers (UMIs) using NEXTflex Rapid Directional qRNA-seq kit (Bioo Scientific, 5130-02D). Libraries were amplified on a LightCycler real-time quantitative PCR machine (Roche) by spiking in SYBR Gold (Life Technologies, S11494). Mid-logarithm amplified libraries were collected and purified using AMPure XP beads (Agencourt, A63881). Libraries were quality controlled by TapeStation (Agilent) and quantitative PCR (KAPA Biosystems).

#### Transduction and Generation of Stable Cell Lines

**[0074]** For individual PiggyBac transductions, cells 500,000 to 800,000 cells were nucleofected using Nucleofector P3 solution (Lonza, V4XP-3032) using the Nucleofector X-Unit. Puromycin-resistant cells were selected and expanded.

**1.-51.** (canceled)

**52.** A method of inducing differentiation of a population of induced pluripotent stem cells, comprising:

delivering to the population of induced pluripotent stem cells:

- (1) a transcription factor;
- (2) a nucleic acid molecule comprising an open reading frame encoding the transcription factor; or

- (3) an activator of transcription of a nucleic acid molecule comprising the open reading frame encoding the transcription factor, or any combination of (1)-(3);

wherein the transcription factor is enriched for one or more DNA-binding domains or depleted for one or more zinc finger domains, one or more Küppel-associated box domains, or both, wherein the one or more DNA-binding domains, the one or more zinc finger domains, the one or more Küppel-associated box domains, or any combination thereof, is enriched or depleted at a level sufficient for the transcription factor to bind a DNA molecule in a sequence-specific manner, and wherein a frequency of normalized reads sequenced of the transcription factor in a differentiated cell is greater than a frequency of normalized reads sequenced of the transcription factor in an undifferentiated cell, thereby differentiating the population of induced pluripotent stem cells.

**53.** The method of claim **52**, wherein the population of induced pluripotent stem cells differentiate in stem cell media. **cm 54.** The method of claim **52**, wherein the population of induced pluripotent stem cells differentiate without embryoid body formation.

**55.** The method of claim **52**, wherein the population of induced pluripotent stem cells differentiate in stem cell media without additional growth factors or mechanical manipulations.

**56.** The method of claim **52**, wherein the transcription factor is not enzymatic and does not form part of a core initiation complex when bound to the DNA molecule.

**57.** The method of claim **52**, wherein <10% of the population of induced pluripotent stem cells express stemness or pluripotency markers, wherein the stemness or pluripotency markers comprise NANOG+, OCT4+, SOX2+, or any combination thereof.

**58.** The method of claim **52**, wherein at least 90% of the population of induced pluripotent stem cells that differentiate lose stem cell identity, wherein the loss in stem cell identity comprises loss of NANOG+, OCT4+, SOX2+, or any combination thereof.

**59.** The method of claim **52**, wherein the one or more DNA-binding domains comprise helix-loop-helix DNA-binding domains.

**60.** The method of claim **52**, wherein the population of induced pluripotent stem cells are human induced pluripotent stem cells.

**61.** The method of claim **52**, wherein the population of induced pluripotent stem cells differentiate in less than around 6 days following transcription factor induction.

**62.** The method of claim **52**, wherein the population of induced pluripotent stem cells differentiate in less than around 4 days following transcription factor induction.

**63.** The method of claim **52**, further comprising using a plurality of a first molecule and a plurality of a second molecule to induce differentiation of a plurality of pluripotent stem cells induces faster differentiation of the plurality of pluripotent stem cells than using a plurality of a first molecule or a plurality of a second molecule alone to induce differentiation of the plurality of pluripotent stem cells.

**64.** The method of claim **63**, wherein the plurality of the first molecule comprises one or more nucleic acid molecules comprising an open reading frame encoding a transcription factor.

**65.** The method of claim **52**, wherein a combination of the transcription factor, the nucleic acid molecule comprising an open reading frame encoding the transcription factor, the activator of transcription of the nucleic acid molecule comprising the open reading frame encoding the transcription factor, are delivered to the population of induced pluripotent stem cells.

**66.** The method of claim **52**, wherein the expression level of a keratin sulfate cell surface antigen in the population of induced pluripotent stem cell is used to determine a proportion of differentiated cell populations from undifferentiated cell populations, wherein the keratin sulfate cell surface antigen is TRA-1-60.

**67.** The method of claim **52**, wherein the transcription factor that differentiates the population of induced pluripotent stem cells is highly enriched in TRA-1-60<sup>low</sup> cell populations compared to a TRA-1-60<sup>high</sup> cell population.

**68.** The method of claim **52**, further comprising transplanting the population of induced pluripotent stem cells into a patient.

**69.** The method of claim **68**, wherein the population of induced pluripotent stem cells are derived from somatic cells of the patient.

**70.** The method of claim **68**, further comprising contacting the differentiated population of induced pluripotent stem cells with a test substance and observing a change in the differentiated population of induced pluripotent stem cells induced by the test substance.

**71.** The method of claim **52**, further comprising inducing differentiation of the population of pluripotent stem cells into a predetermined cell type.

\* \* \* \* \*